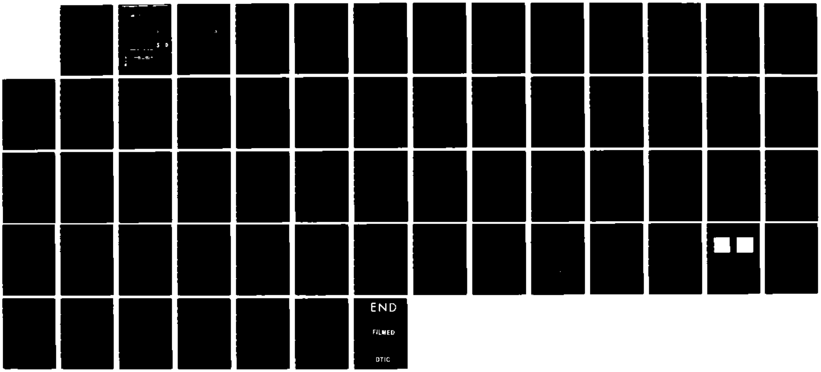
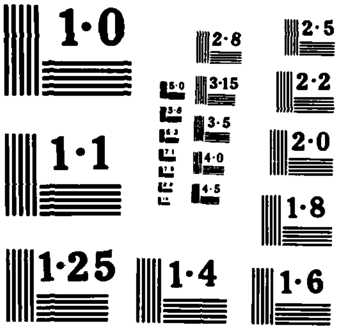
AD-A158 089 BINOCULA IMAGE FLOWS: STEPS TOWARD STEREO - MOTION 1/1
FUSION(U) MARYLAND UNIV COLLEGE PARK CENTER FOR
AUTOMATION RESEARCH A M WAXMAN ET AL. MAY 85
UNCLASSIFIED CAR-TR-119 DAAK70-83-K-0018 F/G 6/16 NL





NATIONAL BUREAU OF STANDARDS
MICROCOPY RESOLUTION TEST CHART

CAR-TR-119,
CS-TR-1494

DAAK70-83-K-0018
May 1985

BINOCULAR IMAGE FLOWS: STEPS TOWARD STEREO - MOTION FUSION

Allen M. Waxman
Computer Vision Laboratory
Center for Automation Research
University of Maryland
College Park, MD 20742

James H. Duncan
Flow Research Company
Silver Spring, MD 20910

COMPUTER VISION LABORATORY

CENTER FOR AUTOMATION RESEARCH

UNIVERSITY OF MARYLAND
COLLEGE PARK, MARYLAND
20742

CAR-TR-119,
CS-TR-1494

DAAK70-83-K-0018
May 1985

BINOCULAR IMAGE FLOWS: STEPS TOWARD STEREO - MOTION FUSION

Allen M. Waxman
Computer Vision Laboratory
Center for Automation Research
University of Maryland
College Park, MD 20742

James H. Duncan
Flow Research Company
Silver Spring, MD 20910

ABSTRACT

The analyses of visual data by stereo and motion modules have typically been treated as separate, parallel processes which both feed a common viewer-centered 2.5-D sketch of the scene. When acting separately, stereo and motion analyses are subject to certain inherent difficulties; stereo must resolve a combinatorial correspondence problem and is further complicated by the presence of occluding boundaries; motion analysis involves the solution of nonlinear equations and yields a 3-D interpretation specified up to an undetermined scale factor. ~~A~~ new module is described here which unifies stereo and motion analysis in a manner in which each helps to overcome the other's shortcomings. One important result is a *correlation between relative image flow (i.e., binocular difference flow) and stereo disparity;* it points to the importance of the ratio δ/δ, rate of change of disparity δ to disparity δ, and its possible role in establishing stereo correspondence. Our formulation may reflect the human perception channel probed by Regan and Beverley (1979).

The support (of A.M.W.) by the Defense Advanced Research Projects Agency and the U.S. Army Night Vision Laboratory under Contract DAAK70-83-K-0018 (DARPA Order 3206) is gratefully acknowledged, as is support (of J.H.D.) by the Department of Energy under Contract DE-AC03-84ER8[illegible].

1. INTRODUCTION

In decomposing the visual information processing task into several stages, it is the intermediate level which is responsible for the recovery of surface shapes in a scene (Marr 1982). It is often described as a set of "shape from" modules which, acting independently and in parallel, feed a viewer centered "2.5-D sketch" of the visual field. Two of the most commonly studied and closely related modules are *shape from stereo* (Koenderink and van Doorn 1976; Marr and Poggio 1979; Mayhew and Frisby 1981; Prazdny 1984; Pollard et al. 1985; Eastman and Waxman 1985) and *shape from monocular motion* (Koenderink and van Doorn 1975; Ullman 1979; Prazdny 1980; Longuet-Higgins and Prazdny 1980; Longuet-Higgins 1981; Tsai and Huang 1981a,b; Waxman and Ullman 1983; Waxman 1984; Waxman and Wohn 1984; Wohn and Waxman 1985a,b; Subbarao and Waxman 1985; Buxton et al. 1984). However, when acting independently, each of these processes suffers from certain inherent difficulties; stereo is faced with a combinatorial correspondence problem plagued by the presence of occluding boundaries (Grimson 1981; Poggio and Poggio 1984), while motion analysis involves the solution of nonlinear equations and leaves the 3-D interpretation specified up to an arbitrary scale factor (Waxman and Ullman 1983). There is evidence, however, for a separate channel of human visual processing in which stereo and motion analyses may come together much earlier than at the 2.5-D sketch. We formulate here a theory of time-varying stereo in the context of "binocular image flows," where stereo and motion work closely in order to overcome each other's shortcomings. Central to our approach is the notion of *relative* *flow* (or "binocular difference flow"), representing the difference between image velocities of a feature as seen in the left and right images separately. Neural organizations which perform this "computation" have already been proposed (Regan and Beverley 1979).

The fusion of stereo and motion into a single module has been considered recently by others as well. Richards (1983) demonstrated recovery of structure from orthographic stereo and motion without knowledge of the fixation distance. Jenkin (1984) considered a stereo matching process driven by the 3-D interpretation of feature point velocities. Waxman and Sinha (1984) proposed a "dynamic stereo" technique based upon the relative flow derived from two cameras in known relative motion, valid in the limit of negligible disparity. The question of image motion aiding stereo in *the matching process* was noted by Poggio and Poggio (1984); and as will be shown below, a correlation between binocular difference flow and disparity may support this possibility.

We suggest a decomposition of our stereo-motion module into five steps which begins where low-level vision ends, i.e., it follows the stage of edge and point feature extraction (and tracking over time) in the left and right images separately.

Step 1: Monocular image flow recovery and flow segmentation of the separate left and right image sequences utilizing the *Velocity Functional Method* (Waxman and Wohn 1984) and *overlap compatibility* (Waxman 1984; Wohn and Waxman 1985b). This procedure allows gross correspondence to be established between analytic flow regions in the left and right images. It also reveals the depth and orientation discontinuities that often plague stereo matching and surface reconstruction algorithms.

Step 2: Establishing correspondence between (previously unmatched) left and right image features according to a correlation between binocular difference flow and stereo disparity. This process can be implemented in parallel over the binocular field of view in the context of "local support" within neighborhoods (Prazdny 1984; Pollard et al. 1985; Eastman and Waxman 1985). This correlation points to the importance of the ratio $\dot{\delta}/\delta$, rate of change of disparity $\dot{\delta}$ to disparity δ. A "rigidity assumption" for independently moving objects in the scene also enters here.

Step 3: Use of disparity functionals defined in overlapping neighborhoods to recover smooth surface structure between the discontinuities detected from the monocular flow analyses (Koenderink and van Doorn 1976; Eastman and Waxman 1985).

Step 4: Recovery of rigid body space motions corresponding to separate analytic flow regions utilizing the determined surface structure and either monocular image flow (or a cyclopean image flow). Separate surface patches can then be grouped into rigid objects sharing the same space motions. This process entails solving only linear equations as a measure of its complexity.

Step 5: Use of the separate image flows to track features and discontinuities over time. This allows refinement of disparity estimates to "sub-pixel" accuracy by temporal interpolation. It also allows the *matching process to focus attention* onto areas where new image features will be unveiled and old ones will disappear, i.e., *at the discontinuities and periphery of the field of view.*

This last step suggests that, in the analysis of a time-varying stereo sequence, once an initial correspondence has been determined between left and right images, it is not necessary to establish correspondence anew for the entire image pair at subsequent times. Most of the image features merely flow to new locations which can be predicted. Matching need only be performed on new features which enter the visible field from the periphery and from behind occluding boundaries.

In this paper we formulate several of these steps toward stereo-motion fusion. Section 2 reviews the basic monocular image flow relations for rigid bodies in motion. The importance of locally second-order flows and boundaries of analyticity (i.e., weak and strong flow discontinuities) is stressed as it is important for the binocular flow analysis that follows. In Section 3 we develop the theory of binocular image flows in the context of a parallel stereo configuration, imaging a scene of rigid objects in motion. A correlation is derived between relative flow (binocular difference flow) and stereo disparity, laying the basis for a new kind of matching procedure. This leads us to speculate on the class of "head motions" that are most discerning in light of this correlation. Other relations between monocular flow and binocular flow are obtained as well. In Section 4 we utilize an experimental data set for a short stereo sequence to obtain the measured binocular image flows at one time instant. These flows are then filtered using the Velocity Functional Method, and a flow segmentation is derived in order to detect depth and orientation discontinuities in the scene. This data is then used to confirm the correlation between binocular flow and disparity developed earlier. Section 5 describes two ways that this binocular difference flow-disparity relation may be implemented in order to establish correspondence in the context of "local support." The ability to combine different matching criteria is considered as well. We conclude in Section 6 with a discussion of what remains to be done in the construction of a complete stereo-motion fusion module.

2. MONOCULAR IMAGE FLOWS

Investigations into the recovery of 3-D structure and motion from time-varying monocular imagery have proceeded along two rather distinct paths. One approach has been concerned with the motion of discrete points moving rigidly in space (Ullman 1979; Prazdny 1980; Longuet-Higgins 1981; Tsai and Huang 1981a,b; Adiv 1984). The resulting 3-D interpretation is in the form of rigid body motion parameters and relative depth of points in space. The second approach treats the image flow field as a whole (Koenderink and van Doorn 1976; Longuet-Higgins and Prazdny 1980; Waxman and Ullman 1983; Wohn 1984; Waxman and Wohn 1985) in an attempt to recover the rigid body motion parameters and surface descriptions (slopes and curvatures) of entire surface patches. Recently, work has begun on the 3-D recovery of structure from non-rigid body motions (Ullman 1983; Koenderink, private communication). Our formulation of binocular image flows will follow the continuous field approach developed for monocular flows generated by textured objects in rigid body motion (Waxman and Ullman 1983, Waxman 1984, Waxman and Wohn 1984, Wohn 1984).

We consider a scene as comprised of objects in independent rigid body motion with respect to the observer. The individual objects are imagined as decomposed into surface patches visible to the observer, and these surface patches in space project into neighborhoods in the image. It is actually the surface texture and shading which is observed under perspective projection in the image. Due to the relative motion between object and observer, the projected texture undergoes deformations which reflect the image flow field. The theory of monocular image flows, developed by Waxman and collaborators (cf. References), provides techniques for the recovery of flow fields and deformation parameters from evolving contours, edge fragments and feature points in the imagery, and for recovery of 3-D surface structure and rigid body motion from these deformations. As these ideas provide the starting point for binocular flow analysis, they are reviewed in more detail here.

2.1 Image Velocity Relations

As a textured, rigid object moves through space, the evolving image sequence registered by a monocular observer (e.g. a moving pin-hole camera) contains information in the form of an image flow field. This image flow is determined by the relative rigid body motion between object and observer, as well as the structure of the object's surface visible to the observer. Derivation of this flow field follows that of Waxman and Ullman (1983).

We attribute the relative rigid body motion to an observer represented by the spatial coordinate system (X, Y, Z) in Figure 1. The origin of this system is located at the vertex of perspective projection, and the Z-axis is directed along the center of the instantaneous field of view. The instantaneous rigid body motion of this coordinate system is specified in terms of the translational velocity $V = (V_X, V_Y, V_Z)$ of its origin and its rotational velocity $\Omega = (\Omega_X, \Omega_Y, \Omega_Z)$. The 2-D image sequence is created by the perspective projection of the object onto a planar screen oriented normal to the Z-axis. The origin of the image coordinate system (x, y) on the screen is located in space at $(X, Y, Z) = (0, 0, 1)$; that is, the image is reinverted and scaled to a focal length of unity.

Due to the observer's motion, a point P in space (located by position vector $\boldsymbol{R}$) moves with a relative velocity $\boldsymbol{U} = -(\boldsymbol{V} + \boldsymbol{\Omega} \times \boldsymbol{R})$. At each instant, point P projects onto the screen as point p with coordinates $(x, y) = (X / Z, Y / Z)$. The corresponding image velocities of point p are $(v_x, v_y) = (\dot{x}, \dot{y})$, obtained by differentiating the image coordinates with respect to time and utilizing the components of $\boldsymbol{U}$ for the time derivatives of the spatial coordinates of P. The result is

$$v_x = \left\{ x \frac{V_Z}{Z} - \frac{V_X}{Z} \right\} + [xy\, \Omega_X - (1 + x^2)\, \Omega_Y + y\, \Omega_Z] , \qquad (1a)$$

$$v_y = \left\{ y \frac{V_Z}{Z} - \frac{V_Y}{Z} \right\} + [(1 + y^2)\, \Omega_X - xy\, \Omega_Y - x\, \Omega_Z] . \qquad (1b)$$

These equations define an instantaneous image flow field, assigning a unique 2-D image velocity $\boldsymbol{v}$ to each direction (x, y) in the observer's field of view. For the moment, we shall consider only a single surface patch of some object in the field of view. A small but finite surface patch may be locally approximated by a quadric surface in space as described by six parameters: two slopes, three curvatures and an overall distance scale. If the surface patch is described in this viewer-centered spatial coordinate system by $Z = \varsigma(X, Y)$, then it is straightforward to find the corresponding local representation $Z = Z(x, y)$ as a second-order polynomial in terms of image coordinates. Of these six surface parameters, only five can be recovered directly from the image flow field; *the overall scale factor is lost as it always appears in ratio with the translational velocity V* (Waxman and Ullman 1983). Moreover, the remaining five surface parameters appear in product with the translational space motion. The *kinematic analysis* developed by Waxman and Ullman (1983) leads to a set of twelve algebraic equ..tions relating this 3-D structure and motion to derivatives (through second order) of the image flow. *Recovery of the 3-D information requires solution of nonlinear equations.*

2.2 Second-Order Image Flows

In the recovery of surface structure and 3-D motion from image flow, it is sufficient to describe an image flow as a locally second-order flow field. This has implications with regard to the surfaces which generate the flow itself. For example, a planar surface patch $Z = Z_0 + pX + qY$, may be described exactly as $Z = Z_0(1 - px - qy)^{-1}$ in image coordinates. Substitution into the velocity equations above yields expressions in the form of second-order polynomials. For planar surfaces, such second-order flows are *globally valid.* On the other hand, quadric surfaces generate flows which are not simple polynomials in the image coordinates. However, they may be *locally approximated* as second-order flows. The coefficients of this second-order flow then determine the slopes and (scaled) curvatures of the quadric surface patch as well as its (scaled) space motion. In this context, a complex surface is viewed as a composite of overlapping planar and quadric patches. The image flow associated with a smooth surface is, therefore, a slowly varying (in terms of image coordinates) second-order flow defined over a region of the image.

In order to recover the second-order flow approximation for any neighborhood in the image, it is necessary to have a sufficiently dense texture present in that neighborhood. This texture gives rise to extended contours, edge fragments and point features, all of which are convected along and deformed by the local image flow. These features serve to sample components of the flow field; in particular, the contours and edges yield an estimate of the flow in the direction normal to the contours themselves. The *Velocity Functional Method* (Waxman and Wohn 1984) may then be used to recover the local flow from these sampled components.

We model the components of the local velocity field by second-order polynomials; hence, define the partial derivatives of image velocity evaluated at a local origin as

$$\boldsymbol{v}^{(i,j)} \equiv \left. \frac{\partial^{i+j} \boldsymbol{v}}{\partial x^{i} \partial y^{j}} \right|_{0} . \tag{2}$$

Then the components of instantaneous velocity in the neighborhood are described by the two functionals

$$v_x(x, y) = \sum_{i=0}^{2} \sum_{j=0}^{2}{}_{(i+j \leq 2)} v_x^{(i,j)} \frac{x^{i}}{i!} \frac{y^{j}}{j!} , \tag{3a}$$

Note that (20) requires these combined monocular-binocular flow quantities to be linear functional forms in the variables $(x\,,\,y\,,\,\delta\,)$. Once correspondence is established between left and right images (as in *Step 2*), the measured disparities may be locally fit over small analytic neighborhoods to a linear form motivated by (14), thereby determining local surface structure (as described in *Step 3*). Then equations (20) may be used to fit linear forms to the measured flow quantities over analytic regions (in the least-squares sense), and thus determine the absolute rigid body motion parameters $\boldsymbol{V}$ and $\boldsymbol{\Omega}$ for that region. This requires solving only linear equations. (Recall that structure and motion from monocular flow required solution of nonlinear equations.) This corresponds to *Step 4* of the stereo-motion fusion module described in Section 1.

The obvious symmetries displayed by equations (20) suggest that they may be written in vector notation. Corresponding to the 3-D space position vector $\boldsymbol{R} = (X\,,Y\,,\,Z\,)$ we introduce the 3-D image position vector $\boldsymbol{r} \equiv \boldsymbol{R}\,/Z = (x\,,y\,,1)$. The 3-D image velocity is defined as $\boldsymbol{u} \equiv \dot{\boldsymbol{r}} = (v_x\,,\,v_y\,,\,0)$. Then, recalling that $\Delta v_x \equiv \dot{\delta}$, we can rewrite (20) as

$$\boldsymbol{u} - \frac{\dot{\delta}}{\delta}\boldsymbol{r} \;= - \left(\frac{\boldsymbol{V}}{b}\,\delta + \boldsymbol{\Omega}\times\boldsymbol{r}\right) \qquad (21)$$

It is not coincidental that (21) bears a strong reseinblance to the relation for 3-D space velocity of a point induced by an observer's rigid body motion, i.e., $\boldsymbol{U} \equiv \dot{\boldsymbol{R}} = -\,(\boldsymbol{V} + \boldsymbol{\Omega}\times\boldsymbol{R}\,)$. In fact, (21) is exactly this relationship for $\boldsymbol{U}\,/Z$!

matching: the V_Z head motion using the v_x component, and the Ω_Z head motion using the v_y component. In the experiments to be described in Section 4, we have examined the former case while viewing a frontal plane. The possibility of more complex head motions requires further analysis.

3.4 Monocular - Binocular Flow Relations

In addition to the correlation that exists between binocular difference flow and disparity (13), there are some interesting relations between this binocular flow and the monocular flow (as seen on the cyclopean image, say). The cyclopean image velocity of a feature is the average of the corresponding feature velocities in the left and right images for this parallel stereo configuration. Equations (1) can be interpreted as the monocular flow in cyclopean image coordinates, with space motion parameters and depth interpreted accordingly. Relations (13) and (14) are valid in cyclopean coordinates as well. Replacing $1/Z$ by δ/b in (1) and combining with (13), find

$$v_x - x \left(\frac{\Delta v_x}{\delta} \right) = - \Omega_Y + y \Omega_Z - \frac{V_X}{b} \delta \, , \tag{20a}$$

$$v_y - y \left(\frac{\Delta v_x}{\delta} \right) = \Omega_X - x \Omega_Z - \frac{V_Y}{b} \delta \, , \tag{20b}$$

$$- \left(\frac{\Delta v_x}{\delta} \right) = x \Omega_Y - y \Omega_X - \frac{V_Z}{b} \delta \, . \tag{20c}$$

Equations (20) have been written with "measurable" quantities on the left-hand side and unknown motion parameters as coefficients on the right-hand side.

motions for animals and machines.

For a preliminary determination of the "head motions" that are most discriminating for matching purposes, we have examined Equation (19) for each of the six motion components separately while viewing a planar surface sloped in the X-direction only. In particular, we seek the motions that produce a velocity difference field which is least sensitive to noise from the measurement of the individual velocity fields. More precisely, a motion that facilitates matching must have two characteristics. First, the velocity component used for matching must be measurable. For example, for an Ω_Y head motion, the v_y component is $O(xy)$ while the v_x component is $O(1)$. Thus, the v_y component cannot be measured accurately. However, a V_Z head motion produces velocity components of equal magnitude. In order to discriminate correct from incorrect matches, we also require that for potential matches, the error $(\Delta \boldsymbol{v} - \Delta \boldsymbol{v}|_c)/\boldsymbol{v}$ should scale like the percentage error in the disparity. Table I contains the results of the analysis including the x and y components of the image flow velocity for the cyclopean system ($\boldsymbol{v}$), the velocity differences for incorrect ($\Delta \boldsymbol{v}$) and correct ($\Delta \boldsymbol{v}|_c$) matches, and the error in the velocity difference for incorrect matches divided by the velocity in the cyclopean image $(\Delta \boldsymbol{v} - \Delta \boldsymbol{v}|_c)/\boldsymbol{v}$. From Table I we see that this error function is $O(1)$ in four cases listed in the table: the v_x component for V_Z and Ω_X motions, and the v_y components for Ω_Y and Ω_Z motions. Two of these, v_x for Ω_X motions and v_y for Ω_Y motions, will be inaccurate because the particular velocity component is $O(xy)$ compared to its companion velocity component. Thus we are left with two motions that facilitate denote them by Z_l and Z_r, respectively. Then it is a straightforward exercise, following Section 3.1, to derive an expression for the ratio $\Delta v/\delta$ in the case of an incorrect match. In the cyclopean coordinate system we find,

$$\frac{\Delta v}{\delta} = \frac{\Delta v}{\delta}\bigg|_c - \left[\frac{\delta - \delta_c}{\delta}\right]\begin{Bmatrix}(x+\delta/2)\,\Omega_Y \\ y\ \Omega_Y + \Omega_Z\end{Bmatrix}$$
$$+\left(\frac{Z_r - Z_l}{Z_r}\right)\frac{\delta_c}{\delta}\begin{Bmatrix} V_X/b - (x+\delta/2)\ V_Z/b - \frac{1}{2}(x+\delta/2)\,\Omega_Y \\ V_Y/b - yV_Z/b - \frac{1}{2}y\,\Omega_Y - \frac{1}{2}\Omega_Z\end{Bmatrix}, \qquad (19)$$

where the upper/lower expressions in curly brackets refer to the x/y components of the ratio.

There are two sets of terms which cause the ratio $\Delta v/\delta$ to deviate from its correct value when a false match is chosen. The first set is proportional to the deviation from the correct disparity value, and is generated by relative rotations between the objects and the eyes/cameras. The second set is proportional to the depth error, which vanishes for a frontal plane (it is proportional to the X-component of slope times disparity deviation). This second set is generated by a combination of relative translational and rotational motions. If we consider the case when objects in the scene are stationary and all motions are due to the cyclopean coordinate system, then only one particular motion contributes to every term present. This is the motion Ω_Y, corresponding to a rotation about a vertical axis as due to a rotation of the head about the neck. (Perhaps this is why our eyes are aligned perpendicular to our necks!) It is also interesting to note that translation in the direction of gaze V_Z, contributes to both components of the ratio (19) as well. Both Ω_Y and V_Z are quite natural exploratory head

$$\frac{\dot{\delta}(x,y)}{\delta(x,y)} = \frac{V_Z}{b}\,\delta(x,y) + (y\,\Omega_X - x\,\Omega_Y)\,, \tag{18}$$

which is identical with relation (13a). *Thus, this correlation between relative image flows and stereo disparity is, in fact, a relationship between disparity and its rate of change!*

3.3 Disambiguating "Head Motions"

In Section 5 below, we describe how the correlation between relative flow and disparity (13) or (15), may be used in establishing correspondence between left and right images. But the basic idea is that, for a set of hypothetical matches among features in a neighborhood, the measured ratio of relative flow to disparity should be consistent with a known functional form, i.e., a linear form as suggested by (15). However, if this correlation is to be useful in establishing correspondence, it must be capable of disambiguating false matches. By considering the possibility of false matches, we can examine the class of "head motions" (or camera motions) that generate significant deviations from the derived correlation.

Consider, then, the relative image flow between a feature in the left image and some feature in the right image shifted horizontally by an angle δ and lying along the same epipolar line. When these features do in fact match, the shift δ equals the "correct disparity" δ_c. For a correct match the depth values of the left and right features are equal. But for an incorrect match they need not be;

3.2 Interpreting the Correlation

The correlation between relative flow $\Delta \boldsymbol{v}$ and disparity δ, presented in cyclopean coordinates in (13a,b) is simple to interpret. Recall that we are considering only a parallel stereo imaging geometry, hence, the epipolar lines are horizontal (i.e., parallel to the x-axes). Now the relative flow $\Delta \boldsymbol{v}$ represents the rate of separation of a feature in one image, from its match in the other image. It is the rate of change of vector disparity. As a feature and its match must always lie along some epipolar line, its vertical disparity must remain zero in this case. Thus, relation (13b) expresses the fact that a feature and its match must flow perpendicular to epipolars at the same rate in order to lie on a common epipolar. In general, the rate of change of vertical disparity must be such as to keep a feature and its match on an epipolar line.

For our parallel stereo configuration, we may then identify Δv_x with the rate of change of (horizontal) disparity and denote it by $\dot{\delta}$. Returning to expression (14) we have

$$\dot{\delta} = -\frac{b}{Z^2}\dot{Z} = -\delta\frac{\dot{Z}}{Z} . \qquad (16)$$

From $\boldsymbol{U} = -(\boldsymbol{V} + \boldsymbol{\Omega} \times \boldsymbol{R})$ we have $\dot{Z} = -V_Z - \Omega_X Y + \Omega_Y X$, hence,

$$\frac{\dot{Z}}{Z} = -\frac{V_Z}{Z} - (y\,\Omega_X - x\,\Omega_Y) = -\frac{V_Z}{b}\,\delta - (y\,\Omega_X - x\,\Omega_Y) . \qquad (17)$$

Combining (17) with equation (16) yields for $\dot{\delta}/\delta$,

$$\frac{\Delta v_x(x,y;\delta)}{\delta(x,y)} = \frac{1}{b} V_Z \, \delta(x,y) + (y\,\Omega_X - x\,\Omega_Y) \,, \tag{13a}$$

$$\frac{\Delta v_y(x,y;\delta)}{\delta(x,y)} = 0 \,, \tag{13b}$$

with image coordinates and motion parameters corresponding to the cyclopean coordinate system.

If we consider the relative flow in a small enough neighborhood such that the underlying surface patch may be treated as locally planar, then we have a simple expression for the local disparity field,

$$\delta(x,y) \equiv \frac{b}{Z(x,y)} = \frac{b}{Z_0}(1 - px - qy) \,, \tag{14}$$

where Z_0 is the depth to the plane measured along the center of the cyclopean field of view, and p and q are the components of local slope. Substituting (14) for the disparity on the right-hand side of (13) yields the local relative flow to disparity relations,

$$\frac{\Delta v_x(x,y)}{\delta(x,y)} = \frac{V_Z}{Z_0} - \left(\frac{V_Z}{Z_0} p + \Omega_Y \right) x - \left(\frac{V_Z}{Z_0} q - \Omega_X \right) y \tag{15a}$$

$$\frac{\Delta v_y(x,y)}{\delta(x,y)} = 0 \,. \tag{15b}$$

We see that locally, *the relative flow to disparity ratio is a linear function of image coordinates* with coefficients depending on the surface structure and relative motion between object and observer. In Section 5, we shall describe how this correlation between relative flow and disparity can form the basis of a stereo matching procedure.

Equations (9a,b) yield the image velocities of corresponding features in the two cameras/eyes.

Now we define the "relative flow" (or *binocular difference flow*) of features between the left and right images as the difference between the "shifted flow fields", the "shift" being associated with the disparity field;

$$\Delta \boldsymbol{v}\,(x_l\,,\,y_l\,;\,\delta) \equiv \boldsymbol{v}_r\,(x_l + \delta\,[x_l\,,\,y_l\,],\,y_l\,) - \boldsymbol{v}_l\,(x_l\,,\,y_l\,). \tag{10}$$

Upon expanding the coefficient matrices of (9a,b) according to equations (1), forming the relative flow (10) and simplifying yields the following expressions for the components of relative flow;

$$\Delta v_x\,(x_l\,,y_l\,;\,\delta) = \frac{1}{b} V_Z\,\delta^2 + (\,y_l\,\Omega_X\, - x_l\,\Omega_Y\,)\,\delta\;, \tag{11a}$$

$$\Delta v_y\,(x_l\,,y_l\,;\,\delta) = 0\;. \tag{11b}$$

Forming the ratio of relative flow to disparity yields

$$\frac{\Delta v_x\,(x_l\,,y_l\,;\,\delta)}{\delta\,(x_l\,,y_l\,)} = \frac{1}{b} V_Z\,\delta + (y_l\,\Omega_X\, - x_l\,\Omega_Y\,)\;, \tag{12a}$$

$$\frac{\Delta v_y\,(x_l\,,y_l\,;\,\delta)}{\delta\,(x_l\,,y_l\,)} = 0\;. \tag{12b}$$

We shall interpret expressions (12a,b) momentarily. But first note that this ratio of relative flow to disparity is linear in the variables x_l , y_l and δ, with coefficients proportional to the unknown parameters of relative motion. The reader may verify for himself that, when reexpressed in the cyclopean coordinate system (midway between the two cameras/eyes), expressions (12) remain unchanged! Thus, we may suppress the subscript "l" in (12) and write instead, the feature at $(x_l\,,\, y_l\,)$ in the left image. Note that over a particular analytic flow region, the (horizontal) disparity forms an analytic scalar field generated by the smooth depth function $Z_l\,(x_l\,, y_l\,)$. And since the left and right coordinate systems are parallel, the depth function for the corresponding region in the right image may be expressed as

$$\begin{aligned} Z_r\,(x_r\,,\, y_r\,) &= Z_r\,(x_l + \delta\,[x_l\,, y_l\,],\, y_l\,) \\ &= Z_l\,(x_l\,,\, y_l\,). \end{aligned} \tag{7}$$

Let us rewrite the monocular image velocity relations (1) in terms of translation and rotation coefficient matrices,

$$\boldsymbol{v}\,(x\,,y\,) = \frac{1}{Z\,(x\,,y\,)}\; \vec{\vec{T}}\,(x\,,y\,) \cdot \boldsymbol{V} + \vec{\vec{R}}\,(x\,,y\,) \cdot \boldsymbol{\Omega}\;; \tag{8}$$

these 2 × 3 matrices being functions of image coordinates alone with elements easily obtained from relations (1). Now an expression like (8) may be associated with each image in our stereo configuration; the coordinates, motion parameters and depth function are, however, different. In order to relate the left and right image flows for a given region, we shall express both flows in terms of the left coordinate system by using expressions (5,6,7). Thus, the left image flow is given by

$$\boldsymbol{v}_l\;\,(x_l\,,\, y_l\,) = \frac{1}{b}\;\delta\,(x_l\,,\, y_l\,)\;\vec{\vec{T}}\,(x_l\,,\, y_l\,) \cdot \boldsymbol{V}_l + \vec{\vec{R}}\,(x_l\,,\, y_l\,) \cdot \boldsymbol{\Omega}_l\;, \tag{9a}$$

while the right image flow is given by

$$\boldsymbol{v}_r\,(x_l + \delta, y_l\,) = \frac{1}{b}\;\delta\,(x_l\,, y_l\,)\;\vec{\vec{T}}\,(x_l + \delta, y_l\,) \cdot \left\{ \boldsymbol{V}_l - \boldsymbol{\Omega}_l \times\; b\hat{\boldsymbol{i}} \right\} + \vec{\vec{R}}\,(x_l + \delta, y_l\,) \cdot \boldsymbol{\Omega}_l \tag{9b}$$

that the left and right cameras/eyes are in motion with respect to each other when relative motion between object and observer is ascribed to the observer. If according to the left coordinate system the rigid body motion parameters of a region are $(\boldsymbol{V}_l , \boldsymbol{\Omega}_l)$, then in the right coordinate system that same region has motion parameters $(\boldsymbol{V}_r , \boldsymbol{\Omega}_r)$, where

$$\boldsymbol{\Omega}_r = \boldsymbol{\Omega}_l \tag{5a}$$

$$\boldsymbol{V}_r = \boldsymbol{V}_l - \boldsymbol{\Omega}_l \times b\hat{\boldsymbol{i}}, \tag{5b}$$

and $\hat{\boldsymbol{i}}$ is a unit vector in the common x-direction.

Thus, the image flow fields of the two eyes/cameras differ in magnitude as well as distribution (due to stereo disparity). And as both stereo disparity and monocular flow vary inversely with depth, we should not be surprised that binocular flow and disparity are related in a simple way. In fact, we shall see that binocular flow is synonymous with "rate-of-change of disparity."

3.1 Relative Flow - Disparity Relation

Given the parallel stereo configuration, we have the simple case of corresponding features lying along horizontal epipolar lines. Thus, a feature located at position (x_l , y_l) in the left image at some instant of time is located at (x_r , y_r) in the right image, where

$$y_r = y_l , \tag{6a}$$

$$\delta(x_l , y_l) \equiv x_r - x_l = b / Z_l \ (x_l , y_l), \tag{6b}$$

$\delta(x_l , y_l)$ being the angular disparity between right and left image positions of

3. BINOCULAR IMAGE FLOWS

For simplicity, we restrict our analysis to the parallel stereo configuration illustrated in Figure 3. The left and right image planes lie in a common plane with the fixation point located at infinity (i.e., the "eyes" point straight ahead). The left and right coordinates, $(x_l\,,\,y_l)$ and $(x_r\,,\,y_r)$ respectively, have their origins at the centers of their respective fields of view separated by a baseline of magnitude b along the common direction of the x-axes. Each image plane is positioned at a focal length of unity with respect to a pin-hole located at the vertex of projection for each separate camera/eye. *This stereo configuration is assumed to move rigidly with respect to other moving objects in the scene.* No allowance has been made for vergence of the eyes (known or otherwise) in the current formulation.

Consider the monocular flow analysis of *Step 1* already performed separately on the left and right image sequences. The analytic flow regions bounded by flow discontinuities are assumed to be brought into correspondence rather easily. This can be accomplished essentially by matching the flow discontinuities between left and right images. The correspondence is gross, but allows the binocular flow analysis to focus attention on individual regions. Each such region is assumed to correspond to a smooth surface of a rigid body. Thus, we may associate with each region a set of relative rigid body motion parameters. However, for the sake of analysis, if we ascribe the rigid body motion to the "monocular observer", as in Figure 1 and equations (1), then the rigid body motion parameters for a given region are different for the left and right cameras/eyes. This is due to the fact sitates the splitting and merging of neighborhoods in order to localize this discontinuity. The beginnings of a control structure governing the automatic segmentation of flow fields is presented in Section 4 below.

2.4 Monocular Analysis of Binocular Flows

In the case of a binocular image sequence, the monocular flow analysis described above is to be applied to the left and right image sequences separately. But rather than going so far as the 3-D inference from monocular flow (Waxman and Ullman 1983) for each sequence, we consider only the recovery and segmentation of the separate image flows. This segmentation into analytic regions (i.e., regions of slowly varying second-order flow) allows gross correspondence to be established between these regions in the left and right images. It also delineates the depth and orientation discontinuities which often plague stereo matching and surface reconstruction algorithms.

This completes *Step 1* of our stereo-motion fusion module. The reconstructed flow fields for the left and right images are brought together in the stage of "binocular flow analysis" described next.

2.3 Boundaries of Analyticity

From equations (1) it is apparent that the flow field is "functionally analytic" (i.e. twice differentiable) wherever object surfaces $Z(x, y)$ are twice differentiable. The flow is non-analytic at points where Z or its first partials are discontinuous, and where the relative space motion parameters change. Such points occur along occluding boundaries and structural edges where surface orientation changes abruptly (e.g., the edges of a polyhedron). Thus, an image flow field is naturally partitioned into regions of analyticity separated by singular contours (i.e., *boundaries of analyticity*). These analytic regions are, in turn, decomposed into neighborhoods in which the image flow is locally approximated as a second-order flow. It is part of a complete image flow analysis to delineate these boundaries of analyticity so that 3-D interpretations can be assigned to the regions within them. Figure 2 illustrates this partitioning of the image flow field.

In order to detect the presence of a boundary of analyticity in the flow field, we try to "analytically continue" the flow from one neighborhood to the next. This is accomplished by requiring the separate second-order flow approximations determined in each neighborhood to be "compatible" in an overlapping area common to both neighborhoods (Wohn 1984; Wohn and Waxman 1985b). The degree of compatibility between neighboring flow approximations is measured relative to the agreement between the individual approximations and the data from which they are obtained. When neighboring flow approximations are deemed "incompatible," it is assumed that a boundary of analyticity has been crossed. This neces-

$$v_y\ (x\,,\,y\,) = \sum_{\substack{i=0 \\ (i+j\le 2)}}^{2}\ \sum_{j=0}^{2} v_y{}^{(i,j)}\,\frac{x^i}{i\,!}\,\frac{y^j}{j\,!}\ . \tag{3b}$$

Now consider a contour or edge fragment embedded in the neighborhood, along which the normal flow has been measured; let this normal flow be given by $v_n\ (x\,,y\,)$. Also, let the unit normal measured along the contour be given by $\boldsymbol{n}\ (x\,,y\,) = (n_x\,,n_y\,)$. Then, since $v_n\ \equiv \boldsymbol{v}\cdot\boldsymbol{n}$, it follows from (3a,b) that

$$v_n\ (x\,,\,y\,) = \sum_{\substack{i=0 \\ (i+j\le 2)}}^{2}\ \sum_{j=0}^{2}\frac{x^i}{i\,!}\,\frac{y^j}{j\,!}\left\{n_x\,(x\,,y\,)\,v_x{}^{(i,j)} + n_y\,(x\,,y\,)\,v_y{}^{(i,j)}\right\}\cdot \tag{4}$$

Equation (4) relates the normal flow along the contour to the twelve parameters (Taylor coefficients) that characterize the full flow in the neighborhood. For each point along a contour at which normal flow and the unit normal are measured, expression (4) provides another constraint on these twelve coefficients. In principle, twelve measurements along a contour are the minimum required to obtain a set of twelve linear equations for the twelve unknowns. In practice, it is better to use many (perhaps hundreds of) measurements along a single or multiple contours and edges in a neighborhood, and let equation (4) serve as the basis of a least-squares approach for obtaining the set of twelve linear equations. Image velocity measurements at points can easily be incorporated into (4) by choosing $\boldsymbol{n}$ along the direction of point motion. The *Velocity Functional Method* has been extended to incorporate data from multiple frames by considering time-varying flows (Wohn and Waxman 1985b). In this manner one can essentially smooth the flow fields over time, thereby filtering out additional noise.

3.5 Relation to Dynamic Stereo

An earlier attempt to recover depth to moving objects from relative image flows was termed *Dynamic Stereo* by Waxman and Sinha (1984). The approach was valid in the limit of negligible disparity, i.e., lim $(b/Z) \rightarrow 0$, and required the two cameras to translate with respect to one another in order to develop a difference flow field.

From equation (8) we see that negligible disparity implies, at lowest order, that the coefficient matrices $\overleftrightarrow{T}(x,y)$ and $\overleftrightarrow{R}(x,y)$ are the same for both cameras. Then, if the two cameras can translate with respect to each other by a known amount $\Delta \boldsymbol{V}$, while their relative rotation is zero, a difference flow $\Delta \boldsymbol{v}$ results which is independent of the relative object motions in the scene, i.e., $\Delta \boldsymbol{v} = Z^{-1} \, \overleftrightarrow{T}(x,y) \cdot \Delta \boldsymbol{V}$. A known relative camera motion $\Delta \boldsymbol{V}$ and measured relative flow $\Delta \boldsymbol{v}$ allows determination of depth $Z(x,y)$.

Comparing this to the formulation in Section 3.1 of the binocular difference flow-disparity relation, we see that equations (11) are providing us with higher order terms in powers of $(b/Z) = \delta$, the disparity. In our simulation studies with *Dynamic Stereo,* we found that a finite baseline of one-thousandth the depth would perturb the relative flow, resulting in a depth error of about 2%. This perturbation can be accounted for by considering the terms in equations (11).

4. EXPERIMENTS

A limited experimental program was undertaken to demonstrate the feasibility of implementing the first three steps of the stereo-motion module: *Step 1* (flow recovery and segmentation), *Step 2* (establishing correspondence using the binocular difference flow) and, to a limited extent, *Step 3* (recovering surface structure). Binocular image flow fields were obtained using a camera mounted on a robot arm, viewing scenes consisting of white objects covered by black dots. In general, the experiments were successful insofar as they confirmed the potential of *overlap compatibility* for segmentation of laboratory flow data, and verified the binocular difference flow-disparity relations for a particular configuration. Still, much work remains before a fully automa'ic module is realized.

4.1 Apparatus and Procedures

The moving pair of stereo cameras was simulated using a single, black and white, Sony (model DC-37) CCD-camera mounted on an American Robot, MERLIN robot arm. The images were digitized into 480 × 420 pixel arrays using a Grinnell (GMR-27) display processor and memory. The angular field of view was 27.6 × 24.1 degrees (i.e., 996.7 pixels per radian). Throughout this section, all angular measurements are given in units of pixels; time is in units of seconds. Each image flow field was obtained from three frames taken with the camera at three positions, equally spaced in time, on its trajectory. The trajectories and viewing directions were chosen to simulate a pair of cameras in a parallel stereo configuration (cf. Fig. 3). The baseline between cameras was 3.0 inches.

The scenes consisted of white surfaces covered with a distribution of 0.125 inch diameter black dots. From the typical viewing distance of 40 inches the dots appeared in the image with a diameter of 3 pixels. To obtain the position of the dots in each image, individual images were thresholded and centroids of black regions were found according to:

$$x_c = \sum_{i=1}^{N} \frac{x_i}{N}, \qquad y_c = \sum_{i=1}^{N} \frac{y_i}{N}, \tag{22}$$

where (x_i, y_i) are the image coordinates of the N black pixels in each region. The centroids of the dots were tracked for three frames and velocities at the centroids in the central frame in time were computed according to

$$v_x\big|_{x_c(t),\, y_c(t)} = \frac{x_c(t+\Delta t) - x_c(t-\Delta t)}{2\Delta t}, \tag{23a}$$

$$v_y\big|_{x_c(t),\, y_c(t)} = \frac{y_c(t+\Delta t) - y_c(t-\Delta t)}{2\Delta t}, \tag{23b}$$

which is a central-difference accurate to $O(\Delta t^2)$. The routine that tracks the centroids from frame to frame assumes that the distance from the centroid in the second frame, to the corresponding centroid in the first or third frame, is smaller than the distance to any neighboring feature points. In addition, to insure reasonably accurate velocity measurements, the centroid displacements from frame to frame must be 10 or more pixels. This simple approach limits the density of feature points allowed in any one image. We have used images with about 200 feature points for analysis.

4.2 Image Flow Segmentation

We have analyzed the scene shown in Figure 4, which consists of a planar background with two connected planar surfaces in the foreground. The effective camera motions, also shown in the figure, were 0.25 inches/sec in the viewing direction (toward the scene) and 0.25 inches/second in the X-direction (parallel to the scene). At the central frame the cameras were about 40 inches from the foreground surfaces. Pictures of the image flows obtained in this way are shown in Figure 5. Each velocity field consists of about 260 points.

The current segmentation program reveals the potential locations of flow discontinuities, but does not refine them nor link them into global boundaries of analyticity. The program first divides the image into N^2 equal-sized rectangles; in this case, a 5 $\times$ 5 rectangular grid on each 480 $\times$ 420 pixel image. Each rectangle contained an average of about 10 feature points. The velocity data in each rectangle was then fit to a pair of second-order polynomials (cf. equations 3) using a linear least squares approach. The error per point between the data and the second order fit, defined as

$$err = (N \mid v_{avg} \mid)^{-1} \sum_{i=1}^{N} \left| v_x^i|_{poly} - v_x^i|_{meas} \right| + \left| v_y^i|_{poly} - v_y^i|_{meas} \right| \quad (24)$$

was typically 0.02 .

In an attempt to see if the polynomial flow fields from adjacent rectangles were compatible, i.e., belonged to the same analytic flow region, the velocities were compared in overlapping neighborhoods. Specifically, at vertical boundaries between left and right rectangles and at horizontal boundaries between upper and lower rectangles, an overlap compatibility measure (C_v and C_h, respectively) was computed,

$$C_v = \frac{2.0}{(err_r + err_l)} \left[\frac{1}{A_v} \int\int_{A_v} (v_r - v_l)^2 dxdy \right]^{1/2} , \tag{25a}$$

$$C_h = \frac{2.0}{(err_r + err_l)} \left[\frac{1}{A_h} \int\int_{A_h} (v_r - v_l)^2 dxdy \right]^{1/2} , \tag{25b}$$

where the areas A_v and A_h are shown in Figure 6. After computing the compatibility for the original 5 × 5 rectangular grid, the calculations were repeated twice with the grid shifted to the right in each case by one-third the rectangle width (approximately the distance between feature points). The three horizontal grid positions were then repeated with the grid shifted down by one-half the rectangle height. Thus, the overlap error was computed for the boundaries of 6 rectangular grids with 25 rectangles in each grid. A plot of the overlap compatibility function is shown in Figures 7 and 8 for the vertical boundaries of the left and right images, respectively. Similar plots for the horizontal boundaries appear in Figures 9 and 10. Consider the compatibility across vertical boundaries first, Figures 7 and 8. Note that the contours with $C_v = 4$ (i.e., four times the error in fitting the polynomials) do not correspond to any structural feature of the scene. Thus, the noise level appears to be about 4. In Figure 7, both the vertical occluding boundary and the vertical structural edge appear in the contours with compatability errors as high as 10, i.e., 2.5 times the noise level. For the structural edge (i.e., the slope discontinuity) the largest values appear slightly to the right of the feature. In Figure 8 similar contour shapes are seen, but the vertical occluding boundary is only one rectangle width away from the left side of the picture and is therefore not fully revealed by the contours. Note that these contours also indicate, to some extent, the position of the horizontal occluding boundary. This horizontal boundary is seen more clearly in the compatibility of upper-lower pairs of rectangles, Figures 9 and 10. The compatibility function is again typically 8 to 10 at the boundary.

The flow field segmentation results indicate that the overlap compatibility method can sucessfully locate occluding boundaries (i.e., depth discontinuities) and to some extent structural edges (i.e., slope discontinuities) in real data. However, the noise level and resolution of the results need to be improved. It is believed that both of these problems can be remedied by increasing the density of data points in the images. For small numbers of data points in a neighborhood, the residual between the measured data and the polynomial fit does not reach a stable mean. Thus, both the coefficients of the polynomials and the residual change significantly as data points are added or subtracted from the fit. In the present examples, since only 10 data points were used to fit each polynomial, the results were not statistically stable and random errors contributed to both the residuals in adjacent neighborhoods and the velocity difference in the overlap regions. Thus, the noise level in C_v and C_h was high. The resolution (or localization) problem is controlled by the size of the rectangles and the magnitude of the shift in the grid position. In the present case the rectangles were large (1/5 the image size) but still only contained about 10 data points. The smallest meaningful shift in the grid position is the average distance between data points, in this case about 1/15 the image size. Thus, the low feature point density resulted in low resolution. The low density of feature points in the present example was necessitated by the simple method used to find feature point velocities from three successive frames. This will be modified in future work to remedy the present noise and resolution problems.

4.3 Binocular Flow Field Experiments

In this section we describe a preliminary experimental exploration of the binocular flow equations (11). In particular, a V_Z motion was chosen for the camera pair and the equations were verified. It was pointed out in Section 3.3 that the V_Z motion is one of the two single component motions that will allow accurate discrimination between correctly and incorrectly matched features. The experiment used the camera set-up described earlier to simulate a pair of cameras separated by a 3 inch baseline. The cameras viewed a planar surface perpendicular to the viewing direction (i.e., a frontal plane). The velocity fields were obtained with the cameras at 43.5, 45.0 and 46.5 inches from the surface. The velocity fields obtained in this manner are shown in Figure 11. These velocity fields show the usual pattern with a focus of expansion near the center of the image. Due to problems with the camera mount, it was not possible to align the camera viewing direction with the direction of motion to better than 0.5 degrees.

With the simple motion and scenes used here, it was possible to correct for this misalignment. In future experiments a pair of cameras aligned with a specially designed stereo mount will be used to alleviate this problem.

The binocular flow equations (11) were verified by two techniques: one using the individual data points and the other using the polynomial fits to the velocity fields; the space motion being known in both cases here (which is not generally true). Feature matching using the individual data points will be discussed first. Because of the low density of data points and the fact that matches lie along horizontal epipolars, the pointwise matching problem for this example can be done rather easily. Here we present an example of matching points in the left and right images by trying the various combinations. Table II contains the coordinates and velocities of four points in the right and left images with $y = 98.0 \pm 1.5$ pixels. The potential disparities $(x_l - x_r)$, the difference in the v_x, and $V_z \delta^2/b$ are given for each of the possible sixteen combinations of the two sets of four points. There are two constraints on the correct matches besides satisfying equations (11). First, the disparity must be positive. This is a consequence of the relative positions of the cameras. Second, the velocity difference Δv_x must be positive, as can be seen from equation (11a) with positive b and V_Z. Eight of the sixteen combinations have these two properties. Of the surviving eight combinations, only three have nearly equal values of Δv_x and $V_z \delta^2/b$; they correspond to correct matches. These are:

combination	δ	Δv_z	$V_z \delta^2/b$
1r-2l	71.5	3.1	2.6
2r-3l	71.0	2.9	2.5
4r-4l	70.2	3.2	2.5

The average disparity of 71.4 pixels corresponds to a distance of 41.9 inches, close to the correct value of 45 inches. Below we shall see that this error is due to camera misalignment.

We now turn our attention to matching using the velocity fields derived from the polynomial fits. Using these polynomials and the known space motion, it is possible to obtain an expression for δ as a continuous function of image coordinates. For this example, each image has been divided into 16 rectangular regions with dimensions of 86.4 × 94.4 pixels each. Second-order polynomials have been fit to the velocity data in each region. The polynomials have the form

$$\begin{aligned} (v_z)_l &= B_{0l} + B_{1l}x_l + B_{2l}y + B_{3l}x_l^{\,2} + B_{4l}y^2 + B_{5l}x_l y \ , \\ (v_z)_r &= B_{0r} + B_{1r}x_r + B_{2r}y + B_{3r}x_r^{\,2} + B_{4r}y^2 + B_{5r}x_r y \ . \end{aligned} \tag{26}$$

Defining the potential disparity as $\delta = (x_r - x_l)$, the velocity difference can be expressed as

$$\begin{aligned} \Delta v_z &= (B_{0r} - B_{0l}) + (B_{1r} - B_{1l})x_l + (B_{2r} - B_{2l})y + (B_{3r} - B_{3l})x_l^{\,2} \\ &+ (B_{4r} - B_{4l})y^2 + (B_{5r} - B_{5l})x_l y + (B_{1r} + B_{5r}y + 2B_{3r}x_l)\delta + B_{3r}\delta^2 \ . \end{aligned} \tag{27}$$

When δ is chosen correctly, this polynomial expression will equal $V_Z \delta^2/b$. Thus, we obtain a second-order polynomial in δ whose solution gives the correct disparity value:

$$(B_{3r} - V_Z / b)\, \delta^2 + (B_{1r} + B_{5r}\, y + 2B_{3r}\, x_l)\, \delta + P(x_l, y) = 0 , \qquad (28)$$

where P corresponds to the terms in (27) that are independent of δ.

This polynomial in δ has been solved in each of the rectangles at a point 25 pixels to the right of each rectangle's center in the left image (thus, its match in the right image will be to the left of center in the corresponding rectangle). The disparity should be the same everywhere in the image. The result, averaged over the sixteen rectangles is 78.8 pixels with a standard deviation of 4.0. This corresponds to Z = 38.0 ± 2.0 inches, compared to the correct value of 45.0 inches.

The source of the error is the misalignment of the cameras, as can be seen from the velocity fields below. Because the cameras are moving toward a frontal plane, the focus of expansion of the velocity field should be at the center of the image and the velocity components should be anti-symmetric. The velocity component v_x at the center of each rectangle, averaged over the four rectangles in each of the four columns is given below for the left and right images.

Average Horizontal Velocity (v_x)

	x = -177	x=-59	x=59	x=177
$(v_x)_l \mid_{avg}$	-5.7	-1.8	2.2	6.0
$(v_x)_r \mid_{avg}$	-6.2	-2.3	1.6	5.6

Note that, adding 0.3 to the values of the right image and subtracting 0.15 from the values of the left image would leave both velocity distributions nearly anti-symmetric. The deviations from anti-symmetry correspond to camera misalignments of about 0.5 degrees. These corrections can also be applied to the velocity difference calculations by adding 0.45 to the calculated value. The corrected disparity, averaged over the sixteen rectangles as above, is 66.3 ± 4.5 pixels or 45.1 ± 3.0 inches, which is the correct value. Also subtracting 0.45 from the velocity differences for the individual point combinations in Table II brings the Δv_x and $V_z \delta^2 / b$ values into very close agreement at the correct matches.

5. MATCHING VIA LOCAL SUPPORT

In the case of a static stereo pair of images, many algorithms have been suggested for establishing correspondence between features (i.e., edges and points) in the left and right images. Knowledge of the stereo geometry constrains matches to lie along known epipolar lines (horizontal in the case of our parallel configuration). Recently, several algorithms have emerged which are based on the notion of ***local support of disparity*** (Prazdny 1984; Pollard, Mayhew and Frisby 1985; Eastman and Waxman 1985). Prazdny's algorithm attempts to embody the concept of "coherence" in the local disparity distribution, by assigning a weight to each potential match of a feature based on a measure of similarity between that disparity and potential disparities of other nearby features. Pollard et al. have developed a matching algorithm which is driven by "local consistency with a prescribed disparity gradient limit" of unity (selected on the basis of psychophysical experiments). Again, potential matches of features are found, and the potential disparities of nearby points are tested for compliance with the disparity gradient limit. The approach of Eastman and Waxman is based on the notion of "analytic disparity fields" in overlapping neighborhoods. Potential matches between contours (i.e., extended edges) in the left and right images are established. Then, motivated by (14) the implied disparities are fit to a linear functional form (in the least-squares sense) for potentially matched contours in a neighborhood, thereby yielding a locally planar interpretation along with the average residual (measuring goodness of fit). A match is then selected on the basis of minimizing this residual (and so maximizing local support) subject to the disparity gradient (derived from the functional) being less than a limit of unity. Our use of locally analytic disparity fields is, in fact, a mathematical realization of "coherence." All of these "local support" algorithms may be implemented in a local and parallel manner.

For our case of time-varying stereo, we suggest the use of the binocular difference flow-disparity relation (15) to establish correspondence in our neighborhoods. Of course, the static matching algorithms based on disparity alone may be used as well, but here we explore the additional exploitation of flow to drive the matching procedure. We can implement the matching procedure in either of two ways, both of which embody the concept of "local support" for matching a neighborhood.

Upon considering (15b) first, we see that a feature and its corresponding match along the epipolar should have the same image velocity perpendicular to the epipolar. This may seem to establish correspondence directly, however, it is not very selective since the velocities themselves do not vary greatly. The problem is that (15b) does not describe a trend of variation over a neighborhood, though it does constrain the matching. On the other hand, (15a) is well suited for matching with local support. If in a small neighborhood we approximate the underlying surface as planar, then (15a) suggests that $\dot{\delta}/\delta$ is locally a linear function of the cyclopean image coordinates. Thus, we seek local support for the analytic form

$$\frac{\dot{\delta}(x,y)}{\delta(x,y)} = C_o + C_x x + C_y y \ , \tag{29}$$

where the left hand side consists of measurements $\dot{\delta} \equiv \Delta v_x$ and δ for potential matches, and the coefficients C_o, C_x, C_y are determined in the least squares sense. This approach is appropriate for matching whole contours, where the many disparity measurements implied can be used in the least squares procedure. The matches which minimize the average residual are considered as having maximum local support.

Alternatively, one can seek matches which maximize local support in light of Prazdny's (1984) approach. We first establish all potential matches along epipolars and note the value of $\dot{\delta}/\delta$ corresponding to each potential match for each feature. We then consider, for each feature i, each of its neighbors j over some small area around it. Then choose those matches with values of $(\dot{\delta}/\delta)_i$ and $(\dot{\delta}/\delta)_j$ which are closest. As (29) implies that $\dot{\delta}/\delta$ varies linearly with angular separation, this suggests forming the quantity

$$\omega_{ij} \equiv [\ (\dot{\delta}/\delta)_i - (\dot{\delta}/\delta)_j\ ] \ / \ s_{ij} \ , \tag{30}$$

where $s_{ij}{}^2 \equiv (x_i - x_j)^2 + (y_i - y_j)^2$. Pairs of potential matches which support (29) will generate a value for $\omega_{ij} \sim O(C_x, C_y)$, whereas pairs of matches which don't support (29) lead to $\omega_{ij} \sim O\ (C_o/s_{ij}) >> C_x$ or C_y. As ω_{ij} has units of inverse time, we must adopt a local time constant τ_i and consider the dimensionless quantity $\tau_i \omega_{ij}$ as the primary variable measuring similarity. A reasonable choice for τ is $(\dot{\delta}/\delta)_i{}^{-1} \sim O(C_o{}^{-1})$. Hence, we wish to create a support function which is $O(1)$ when $(\tau_i \omega_{ij})^2$ is small, and then drops to zero as $(\tau_i \omega_{ij})^2$

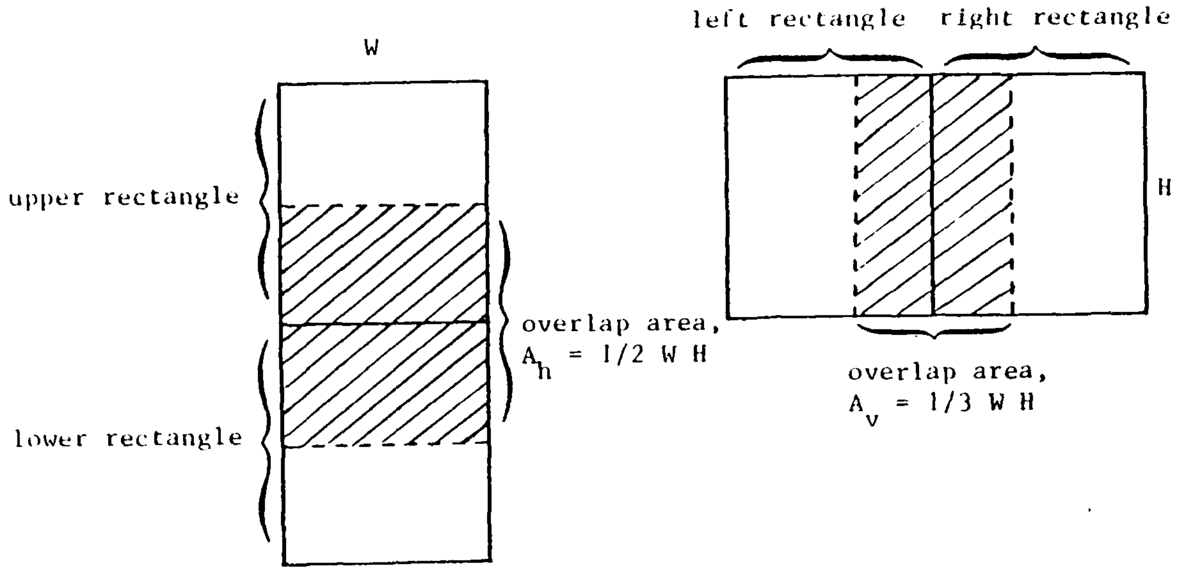
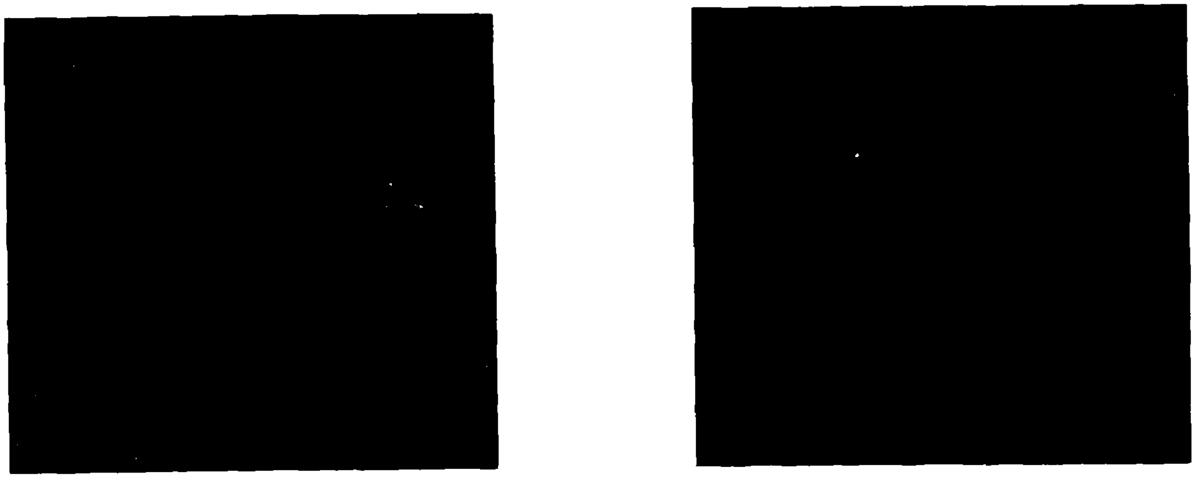


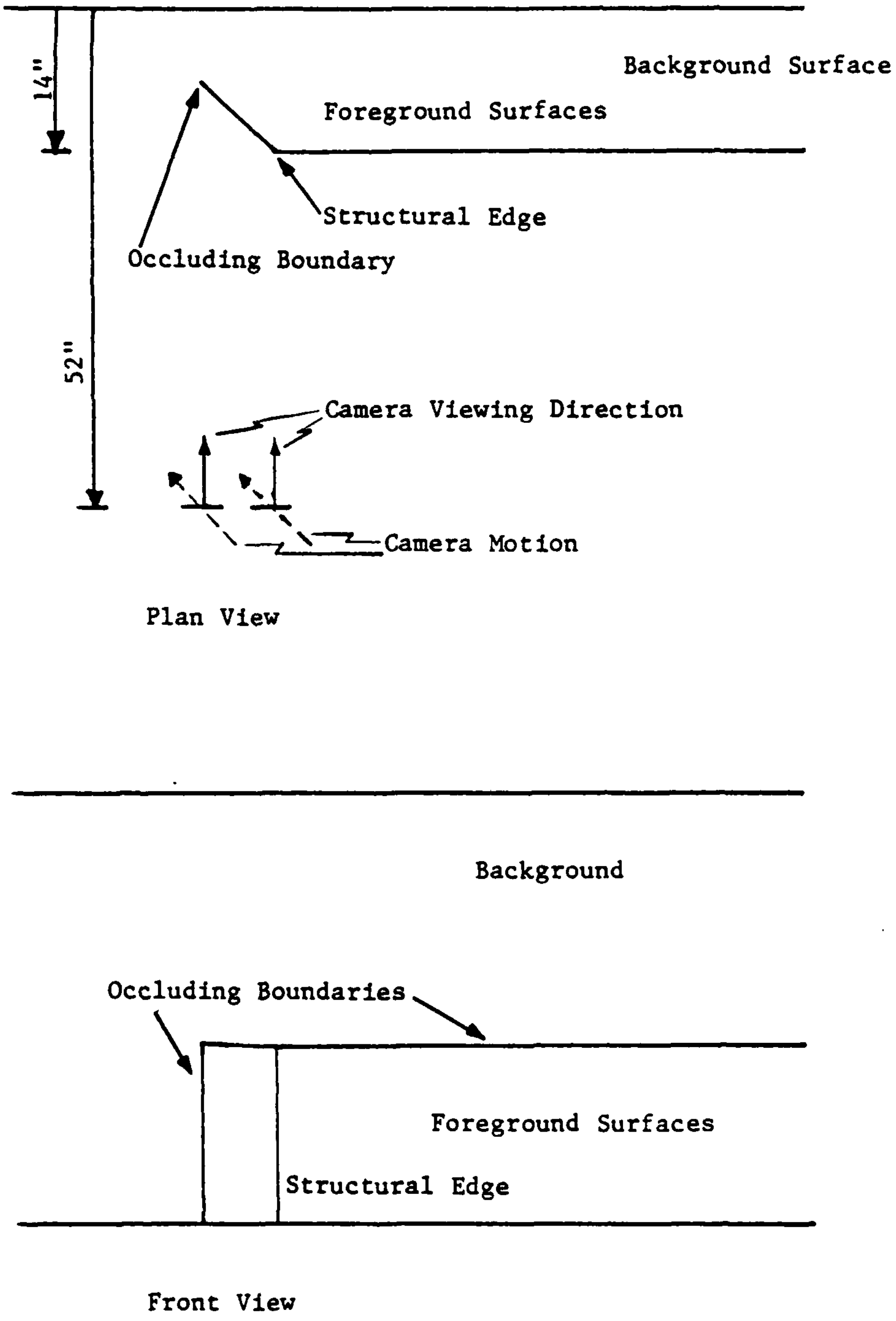
Figure 6 Schematic Showing Overlap Compatibility Areas for Adjacent Rectangular Neighborhoods



Left Image

Right Image

Figure 5 Velocity Fields from Left and Right Images - Segmentation Experiment



Plan View

Front View

Figure 4 Two Views of the Scene Used for the Segmentation Experiments

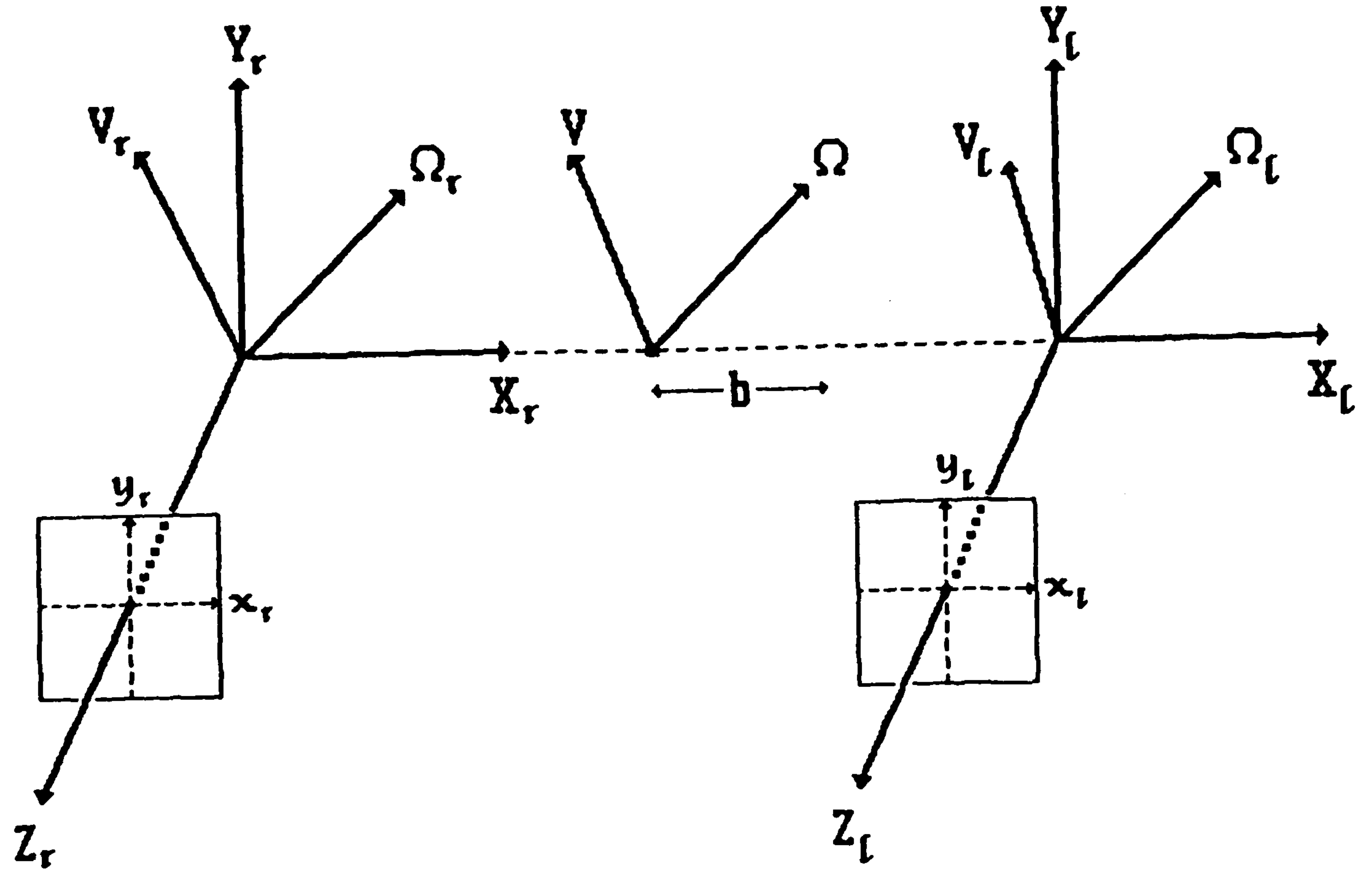


Figure 3 Spatial and Image Coordinates for the Binocular Configuration. Space Motions are Shown for Left, Right and Cyclopean Systems which Move as a Rigid Object.

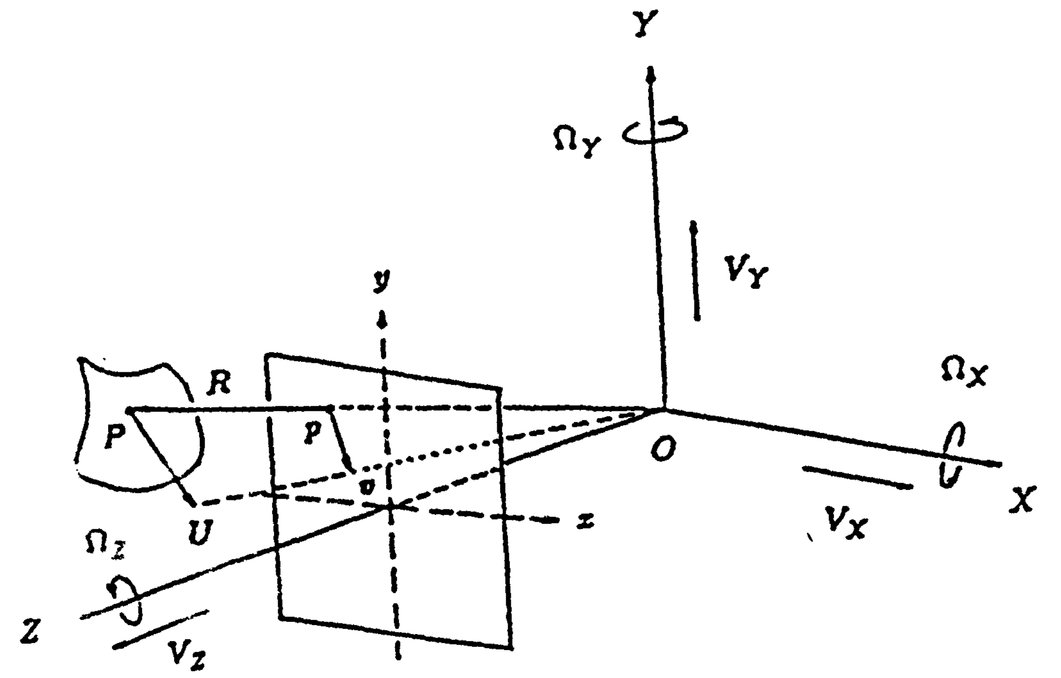


Figure 1 Spatial Coordinates Moving with a Monocular Observer and the Monocular Image Coordinates

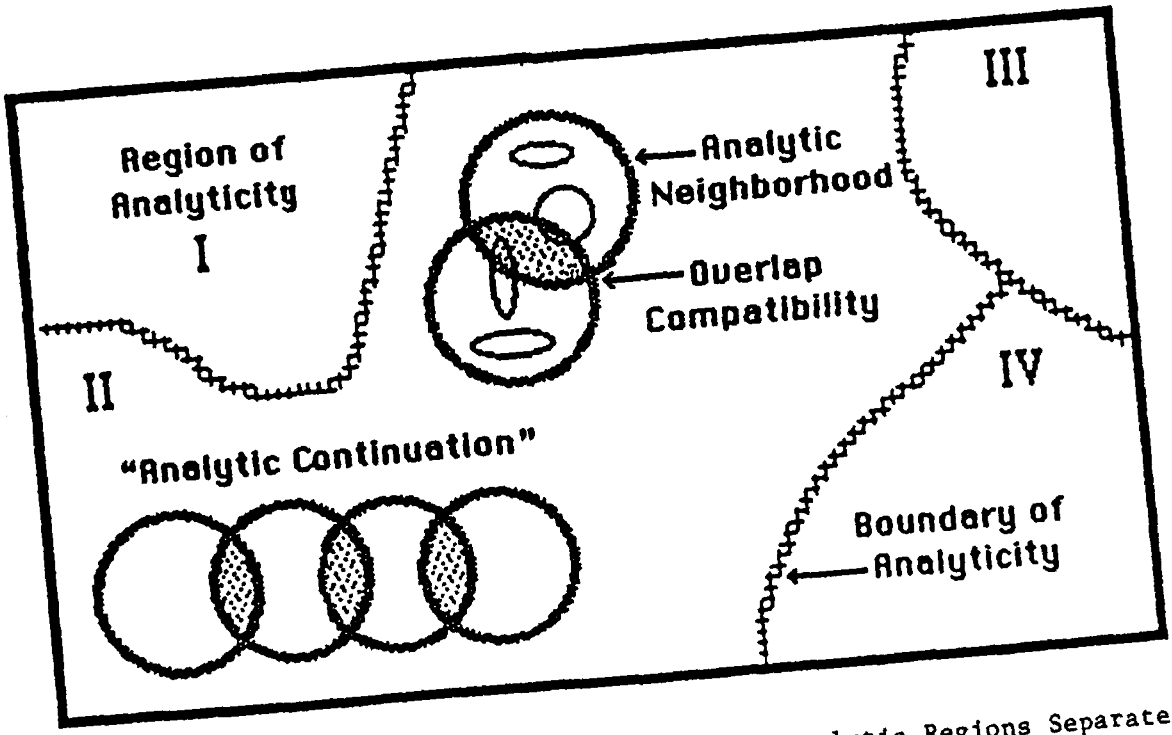


Figure 2 Partitioning the Velocity Image into Analytic Regions Separated by Boundaries of Analyticity. Analytic Regions are Comprised of Overlapping Neighborhoods in which tne Flow Field is Locally Second-Order.

TABLE II
POINTWISE MATCHING WITH
THE BINOCULAR FLOW RELATIONS

Right Image Data Points

point	x	y	v_x	v_y
1r	-44.0	118.7	-1.0	3.8
2r	-122.0	117.7	-3.7	3.7
3r	-163.7	118.0	-5.2	3.7
4r	-0.8	116.8	0.4	3.7

Left Image Data Points

point	x	y	v_x	v_y
1l	190.0	118.3	6.2	3.4
2l	-115.5	118.5	-4.1	3.4
3l	-193.0	117.3	-6.6	3.4
4l	-71.0	116.5	-2.8	3.4

Velocity Difference Data

pair	δ	Δv_x	$V_Z\delta^2/b$	
1r-1l	-234.0	-7.2	27.5	
1r-2l	71.5	3.1	2.6	match
1r-3l	149.0	5.6	11.1	
1r-4l	27.0	1.8	0.4	
2r-1l	-312.0	-9.9	48.8	
2r-2l	-6.5	0.4	0.0	
2r-3l	71.0	2.9	2.5	match
2r-4l	-51.0	-0.9	1.3	
3r-1l	-353.7	-11.4	62.7	
3r-2l	-48.2	-1.1	1.2	
3r-3l	29.3	1.4	0.4	
3r-4l	-92.7	-2.4	4.3	
4r-1l	-190.8	-5.8	18.2	
4r-2l	114.7	4.5	6.6	
4r-3l	192.2	7.0	18.5	
4r-4l	70.2	3.2	2.5	match

TABLE I
DISAMBIGUATING HEAD MOTIONS

	v_x	Δv_x	$\Delta v_x\|_c$	$\frac{\Delta v_x - \Delta v_x\|_c}{v_x}$	v_y	Δv_y	$\Delta v_y\|_c$	$\frac{\Delta v_y - \Delta v_y\|_c}{v_y}$
V_X	$\frac{-V_X}{Z}$	$\delta_c(\delta-\delta_c)\frac{V_z p}{b}$	0	$-p(\delta-\delta_c)$	0	0	0	
V_Y	0	0	0		$\frac{-V_Y}{Z}$	$\frac{V_Y \delta_c}{b}(\delta-\delta_c)p$	0	$-p(\delta-\delta_c)$
V_Z	$\frac{xV_Z}{Z}$	$\delta_c V_Z/b[\delta -$ $p(\delta-\delta_c)(x+\delta/2)]$	$\frac{V_Z \delta_c{}^2}{b}$	$[(\delta-\delta_c)/x]\times$ $[1-p(x+\delta/2)]$	$y\frac{V_Z}{Z}$	$\frac{-V_Z y \delta_c}{b}(\delta-\delta_c)p$	0	$-p(\delta-\delta_c)$
Ω_X	$xy\Omega_X$	$y\Omega_X\delta$	$y\Omega_X\delta_c$	$\frac{\delta-\delta_c}{x}$	$\Omega_X \times$ $(1+y^2)$	0	0	0
Ω_Y	$-(1+x^2)\Omega_Y$	$-\Omega_Y[x\delta+(\delta-\delta_c)\times$ $(x+\delta/2)(1+p)]$	$-x\Omega_Y\delta_c$	$-[(\delta-\delta_c)/(1+x^2)]\times$ $[x+(x+\delta)(1+p\delta_c)]$	$-xy\Omega_Y$	$-y\Omega_Y(\delta-\delta_c)\times$ $(1+p\delta_c/2)$	0	$[(\delta-\delta_c)/x]\times$ $(1+p\delta_c/2)$
Ω_Z	$y\Omega_Z$	0	0	0	$-x\Omega_Z$	$-\Omega_Z(\delta-\delta_c)\times$ $(1+p\delta_c/2)$	0	$[(\delta-\delta_c)/x]\times$ $(1+p\delta_c/2)$

Waxman, A.M. and Ullman, S. 1983 (October). Surface structure and 3-D motion from image flow: A kinematic analysis. Tech. Report 24. College Park, MD: University of Maryland, Center for Automation Researçh. Also see *Int. J. Robotics Research* 4 (3), 1985.

Waxman, A.M. and Wohn, K. 1984 (April). Contour evolution, neighborhood deformation and global image flow: Planar surfaces in motion. Tech. Report 58. College Park, MD: University of Maryland, Center for Automation Research. Also see *Int. J. Robotics Research* 4 (3), 1985.

Waxman, A.M. and Wohn, K. 1985. Contour evolution, neighborhood deformation and image flow: Textured surfaces in motion. *Image Understanding 1985,* (eds.) W. Richards and S. Ullman. Norwood: Ablex Publishing.

Wohn, K. 1984. A contour-based approach to image flow. Ph.D. Thesis, University of Maryland, Department of Computer Science.

Wohn, K. and Waxman, A.M. 1985a. Contour evolution, neighborhood deformation and local image flow: Curved surfaces in motion. Tech. Report in preparation. College Park, MD: University of Maryland, Center for Automation Research.

Wohn, K. and Waxman, A.M. 1985b. The analytic structure of image flows: Deformation and segmentation. Tech. Report in preparation. College Park, MD: University of Maryland, Center for Automation Research.

Poggio, G.F. and Poggio, T. 1984. The analysis of stereopsis. *Ann. Rev. Neurosci.,* 7: 379-412.

Pollard, S.B., Mayhew, J.E.W. and Frisby, J.P. 1985. Disparity gradients and stereo correspondences. Preprint, Dept. Psychology, Sheffield University.

Prazdny, K. 1980. Egomotion and relative depth map from optical flow. *Biol. Cyber.* 36: 87-102.

Prazdny, K. 1984. Detection of binocular disparities. Preprint, Fairchild Laboratory for Artificial Intelligence Research. *Biol. Cyber.* (in press) 1985.

Regan, D. and Beverley, K.I. 1979. Binocular and monocular stimuli for motion in depth: Changing-disparity and changing-size feed the same motion-in-depth stage. *Vision Research* 19: 1331-1342.

Richards, W. 1983. Structure from stereo and motion. A.I. Memo 731. Cambridge, MA: Massachusetts Institute of Technology, Artificial Intelligence Laboratory. See also, *J. Opt. Soc. Amer.* A2: 343-349 (1985).

Subbarao, M. and Waxman, A.M. 1985. On the uniqueness of image flow solutions for planar surfaces in motion. Tech. Report 113. College Park, MD: University of Maryland, Center for Automation Research.

Tsai, R.Y. and Huang, T.S. 1981a. Uniqueness and estimation of 3-D motion parameters of rigid objects with curved surfaces. Report R-921. University of Illinois/Urbana-Champaign Coordinated Science Lab.

Tsai, R.Y. and Huang, T.S. 1981b. Estimating 3-D motion parameters of a rigid planar patch. Report R-922. University of Illinois/Urbana-Champaign Coordinated Science Lab.

Ullman, S. 1979. *The Interpretation of Visual Motion.* Cambridge: MIT Press.

Ullman, S. 1983. Maximizing rigidity: the incremental recovery of structure from rigid and rubbery motion. Memo 721. Cambridge, MA: Massachusetts Institute of Technology Artificial Intelligence Laboratory.

Waxman, A.M. 1984 (April). An image flow paradigm. *Proc. 2nd IEEE Workshop on Computer Vision: Representation and Control,* Annapolis: IEEE, pp. 49-57.

Waxman, A.M. and Sinha, S. 1984 (October). Dynamic Stereo: Passive ranging to moving objects from relative image flows. *Proc. DARPA Image Understanding Workshop,* New Orleans: SAIC, pp. 130-136.

REFERENCES

Adiv, G. 1984 (October). Determining 3-D motion and structure from optical flow generated by several moving objects. *Proc. DARPA Image Understanding Workshop,* New Orleans: SAIC, pp. 113-129.

Burt, P. and Julesz, B. 1980. A disparity gradient limit for binocular fusion. *Science,* 208: 615-617.

Buxton, B.F., Buxton, H., Murray, D.W. and Williams, N.S. 1984. 3-D solutions to the aperture problem. *European Conf. Artificial Intelligence '84.*

Eastman, R. and Waxman, A.M. 1985. Using disparity functionals for stereo correspondence and surface reconstruction. Tech. Report in preparation. College Park, MD: University of Maryland, Center for Automation Research.

Grimson, W.E.L. 1981. *From Images to Surfaces.* Cambridge: M.I.T. Press.

Jenkin, M.R.M. 1984 (September). The stereopsis of time-varying images. Tech. Report RBCV-TR-84-3. Toronto, Canada: University of Toronto, Dept. of Computer Science.

Koenderink, J.J. and van Doorn, A.J. 1975. Invariant properties of the motion parallax field due to the movement of rigid bodies relative to an observer. *Optica Acta* 22: 773-791.

Koenderink, J.J. and van Doorn, A.J. 1976. Geometry of binocular vision and a model for stereopsis. *Biol. Cybernetics* 21: 29-35.

Longuet-Higgins, H.C. 1981. A computer algorithm for reconstructing a scene from two projections. *Nature* 293: 133-135.

Longuet-Higgins, H.C., and K. Prazdny, K. 1980. The interpretation of a moving retinal image. *Proc. Roy. Soc. Lond.* B208: 385-397.

Marr, D. 1982. *Vision.* San Francisco: Freeman.

Marr, D. and Poggio, T. 1979. A computational theory of human stereo vision. *Proc. Roy. Soc. Lond.,* B204: 301-328.

Mayhew, J.E.W. and Frisby, J.P. 1981. Psychophysical and computational studies towards a theory of human stereopsis. *Artificial Intelligence* 17: 349-385.

Still, much work remains to be done before a complete module of this type can be constructed. The control structure for the flow segmentation procedure requires further development. This segmentation procedure should be iterative, with subsequent refinements occurring near detected flow discontinuities. The discontinuities in left and right images must also be matched in order to establish gross correspondence among analytic regions. The binocular difference flow-disparity relation, derived in Section 3, requires further testing in order to insure its validity under more general classes of motion than tried here. It should also be generalized to incorporate vergence effects. The matching techniques described in Section 5 need to be implemented and tested in a variety of cases. The ability to combine evidence in establishing correspondence is an appealing aspect of the approach and needs to be implemented as well.

The possible role of a combined stereo-motion module, such as this one, in the human visual processing task raises some interesting questions. How does the brain utilize disparity estimates and binocular flow-disparity cues in establishing correspondence? Does one take priority over the other, or are they combined? What happens when structure from binocular flow conflicts with structure from static stereo (Mayhew and Frisby, private communication)? Does one percept dominate or do we see illusions? Are there certain kinds of "head motions" preferred for disambiguating false matches? Is there a "gradient limit" effect associated with the coefficients of the linear terms in equation (15a)? Is it possible to fuse a dynamic stereogram which is beyond the static disparity gradient limit of unity? Perhaps psychophysical experiments can resolve some of these questions.

6. CONCLUSIONS

In this paper we have outlined a set of five steps toward the development of a stereo-motion fusion module. The successful development of a complete module of this type has enormous potential for robotics in a dynamic environment. It may also shed some light on the nature of the processing going on in the human visual pathway. In this respect, the work of Regan and Beverley (1979) is most relevant, for their own psychophysical and neurophysiological studies have led them to suggest the existence of neural organizations which may "compute" the binocular difference flow (or relative flow between the eyes) which is so basic to our own theory.

The basic advantages this module offers over static stereo are: monocular detection of the depth and orientation discontinuities (before matching is attempted), use of a correlation between binocular difference flow and disparity to drive the matching process (either independent of, or in conjunction with matching based on disparity alone), the ability to refine disparity estimates to sub-pixel accuracy by considering the smooth orbits of features through the left and right image space-times, and the potential to focus attention of the matching process to the areas where new features enter the field of view. The advantages of this module over structure from monocular motion are: the ability to recover absolute structure and rigid body motions (without scale factor ambiguities), and that only linear equations need be solved to recover rigid body motion parameters.

above. One can require either independent confirmation of a match (both processes running in parallel lead to the same conclusion), or combined evidence of a match based on redundant support (using the product of independent support functions, hence the logical "AND") or complementary support (using the sum of independent support functions, hence the logical "OR"). This method of combining evidence for matching awaits implementation as well.

grows. The function should seek support over only a local neighborhood around feature i. Denoting this function by $W\,(\tau_i\,\omega_{ij})$, we form $\sum_j W\,(\tau_i\,\omega_{ij})$ over the neighborhood and select the match for feature i with value $(\dot{\delta}/\delta)_i$ that generated the largest percentage of the sum; it is most similar to its neighbors in a manner consistent with the linear form (29).

This is essentially Prazdny's algorithm, adapted to the variable $\dot{\delta}/\delta$. Clearly, it is applicable to any variable which can be locally approximated as a linear form, including disparity itself. Such a matching strategy leads naturally to a preference for small gradients in the matching variable. Thus, a kind of "gradient limit" emerges. This is well known for disparity alone in static stereograms (Burt and Julesz 1980). But does such a gradient limit exist for dynamic stereograms? Could fusion be achieved with a dynamic stereogram for which the disparity gradient limit is exceeded?

We have yet to implement our matching strategy and so cannot comment on its possible strengths or weaknesses. But in keeping with *Step 5* of Section 1, we expect that once correspondence is initially established, new features emerging from behind occluding boundaries and the periphery are easily matched. They are entrained into the local disparity field by a spreading of local support from previously matched features in the neighborhood.

Finally, we can consider the possibility of combining multiple matching criteria. For example, disparity δ and the ratio $\dot{\delta}/\delta$ may both be used to establish correspondence, and can both be implemented in the same fashion as outlined

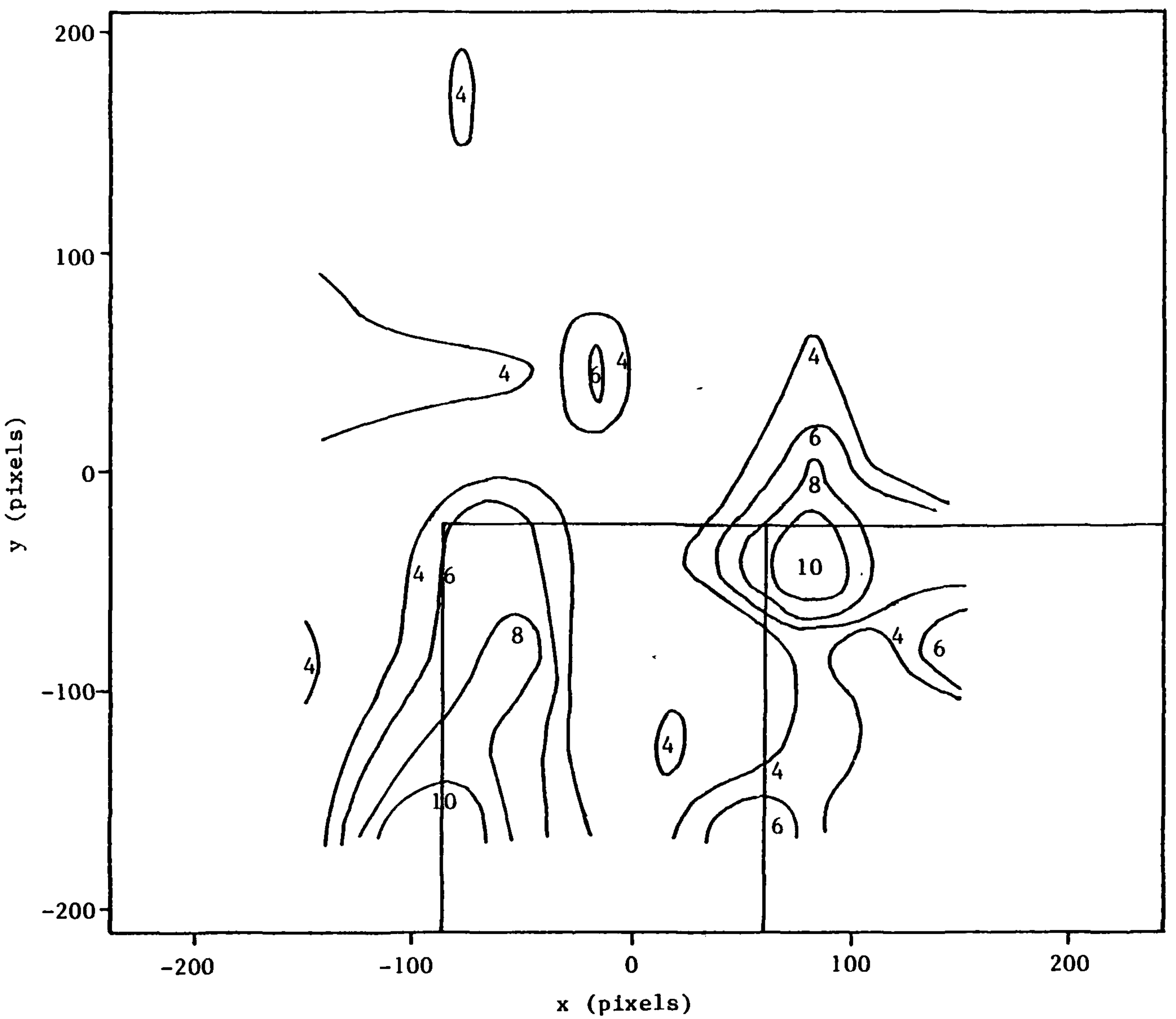


Figure 7 Overlap Compatibility Contours Across Vertical Boundaries, C_v - Left Image

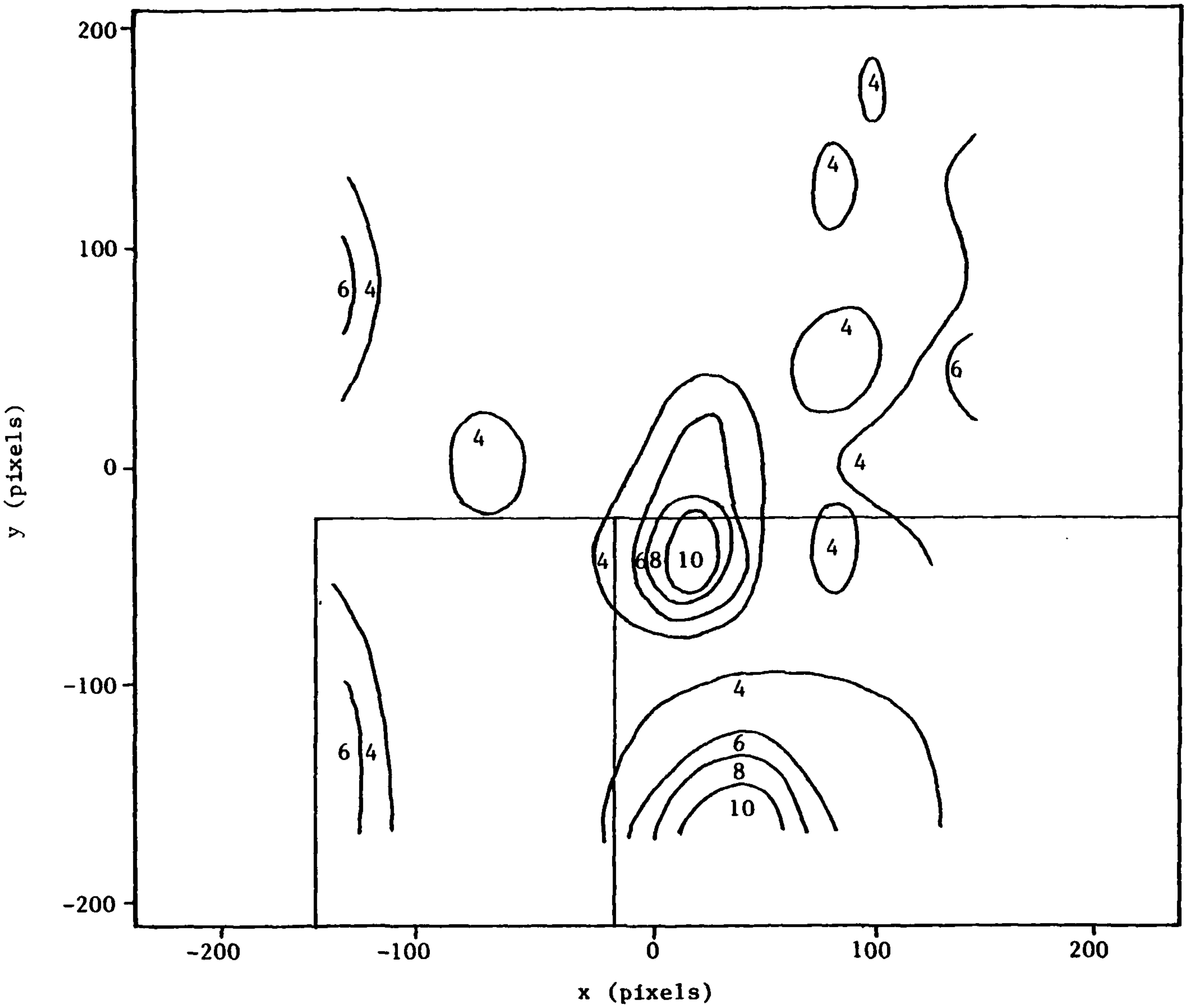


Figure 8 Overlap Compatibility Contours Across Vertical Boundaries, C_v - Right Image

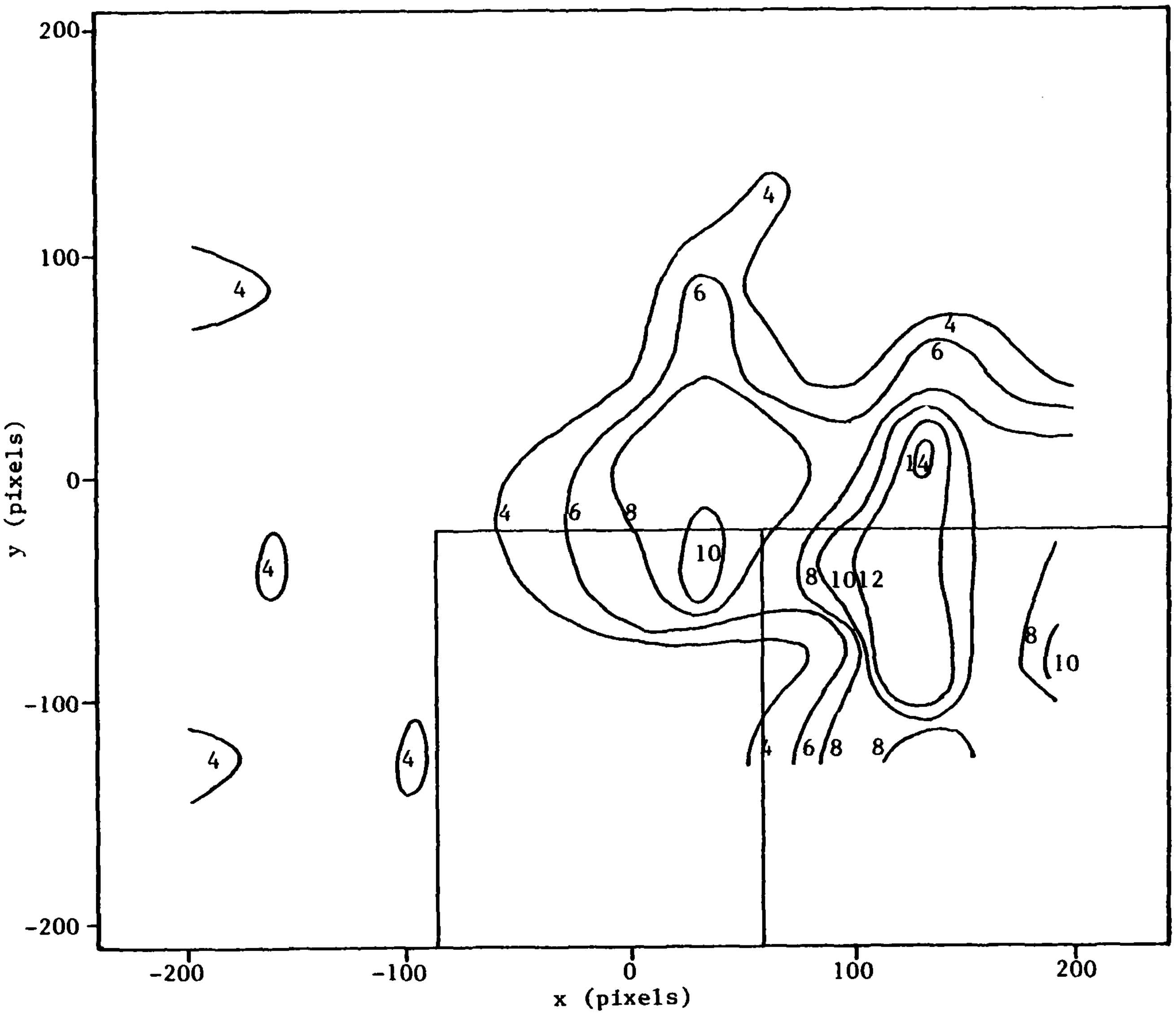


Figure 9 Overlap Compatibility Contours Across Horizontal Boundaries, C_h - Left Image

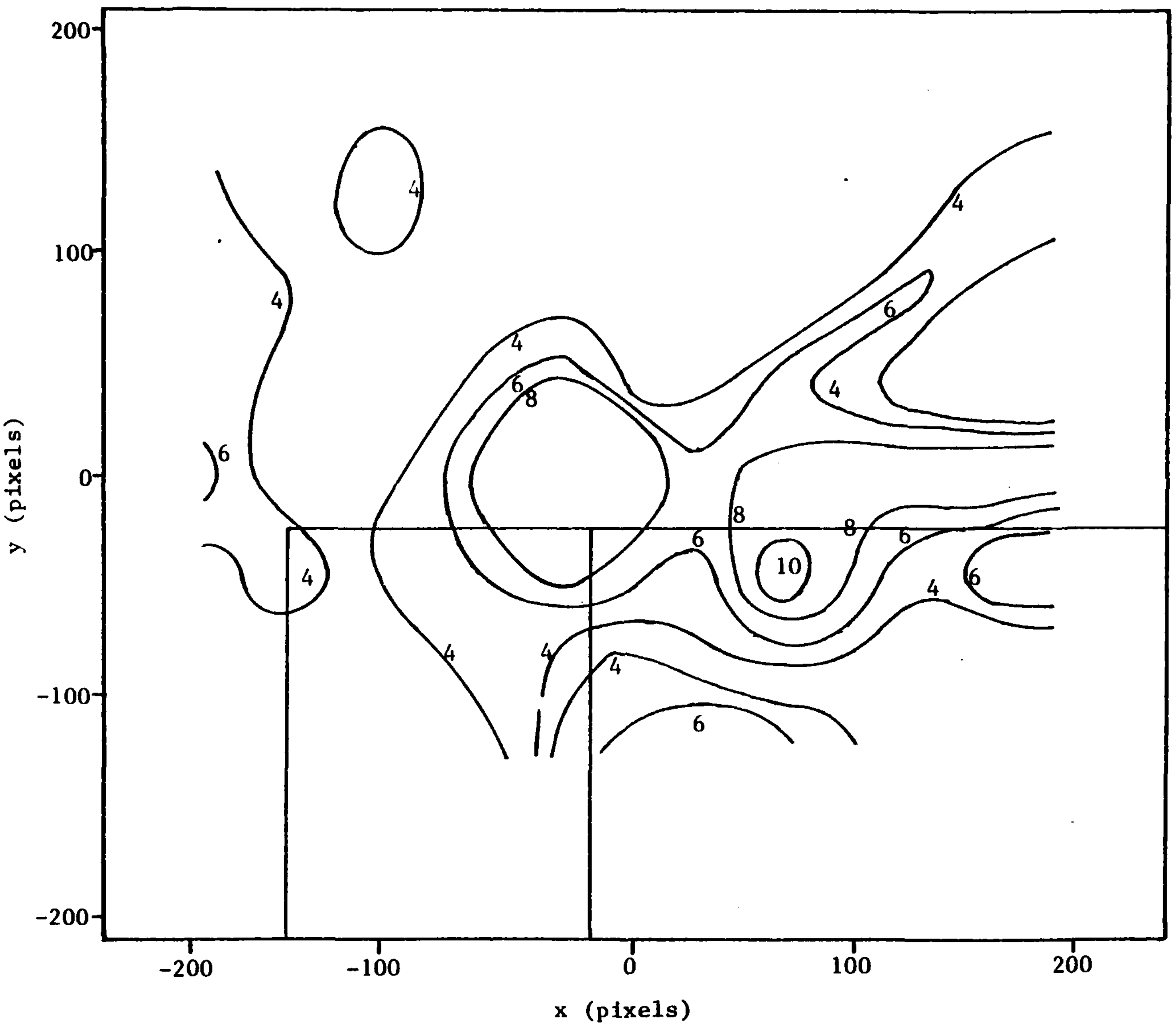
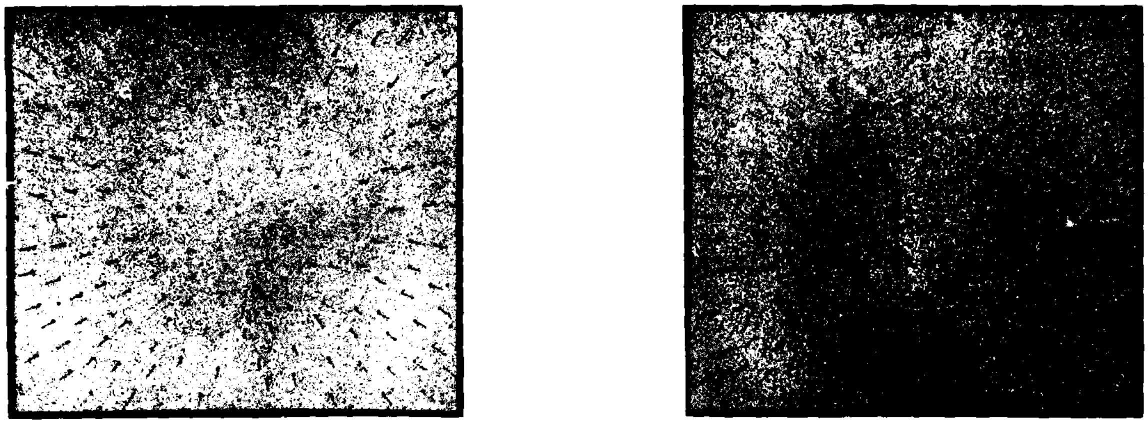


Figure 10 Overlap Compatibility Contours Across Horizontal Boundaries, C_h - Right Image



Left Image

Right Image

Figure 11 Velocity Fields from Left and Right Images - Cameras Moving Toward a Frontal Plane

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION
Unclassified

1b. RESTRICTIVE MARKINGS
N/A

2a. SECURITY CLASSIFICATION AUTHORITY
N/A

2b. DECLASSIFICATION/DOWNGRADING SCHEDULE
N/A

3. DISTRIBUTION/AVAILABILITY OF REPORT
Approved for public release; distribution unlimited

4. PERFORMING ORGANIZATION REPORT NUMBER(S)
CAR-TR-119
CS-TR-1494

5. MONITORING ORGANIZATION REPORT NUMBER(S)
N/A

6a. NAME OF PERFORMING ORGANIZATION
University of Maryland

6b. OFFICE SYMBOL (If applicable)
N/A

7a. NAME OF MONITORING ORGANIZATION
Army Night Vision and Electro-Optics Laboratory

6c. ADDRESS (City, State and ZIP Code)
Center for Automation Research
College Park, MD 20742

7b. ADDRESS (City, State and ZIP Code)
Fort Belvoir, VA 22060

8a. NAME OF FUNDING/SPONSORING ORGANIZATION
Defense Advanced Research Projects Agency

8b. OFFICE SYMBOL (If applicable)
IPTO

9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
DAAK70-83-K-0018

8c. ADDRESS (City, State and ZIP Code)
1400 Wilson Blvd.
Arlington, VA 22209

10. SOURCE OF FUNDING NOS.

PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT NO.

11. TITLE (Include Security Classification)
Bionocular image flows: steps toward stereo-motion fusion

12. PERSONAL AUTHOR(S)
Allen M. Waxman, James H. Duncan

13a. TYPE OF REPORT
Technical

13b. TIME COVERED
FROM ________ TO N/A

14. DATE OF REPORT (Yr., Mo., Day)
May 1985

15. PAGE COUNT
58

16. SUPPLEMENTARY NOTATION

17. COSATI CODES

FIELD	GROUP	SUB. GR.

18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

The analyses of visual data by stereo and motion modules have typically been treated as separate, parallel processes which both feed a common viewer-centered 2.5-D sketch of the scene. When acting separately, stereo and motion analyses are subject to certain inherent difficulties; stereo must resolve a combinatorial correspondence problem and is further complicated by the presence of occluding boundaries, motion analysis involves the solution of nonlinear equations and yields a 3-D interpretation specified up to an undetermined scale factor. A new module is described here which unifies stereo and motion analysis in a manner in which each helps to overcome the other's shortcomings. One important result is a correlation between relative image flow (i.e., binocular difference flow) and stereo disparity; it points to the importance of the ratio $\dot{\delta}/\delta$, rate of change of disparity $\dot{\delta}$ to disparity δ, and its possible role in establishing stereo correspondence. Our formulation may reflect the human perception channel probed by Regan and Beverley (1979).

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT
UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT. □ DTIC USERS □

21. ABSTRACT SECURITY CLASSIFICATION
Unclassified

22a. NAME OF RESPONSIBLE INDIVIDUAL

22b. TELEPHONE NUMBER (Include Area Code)

22c. OFFICE SYMBOL

EDITION OF 1 JAN 73 IS OBSOLETE.

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE